HANJO HAMANN

Evidenzbasierte Jurisprudenz

Grundlagen der Rechtswissenschaft
23

Mohr Siebeck

Grundlagen der Rechtswissenschaft

herausgegeben von

Horst Dreier, Ulrike Müßig und Michael Stolleis

23



Hanjo Hamann

Evidenzbasierte Jurisprudenz

Methoden empirischer Forschung
und ihr Erkenntniswert für das Recht
am Beispiel des Gesellschaftsrechts

Mohr Siebeck

Hanjo Hamann, Studium der Rechtswissenschaft in Heidelberg und Hamburg; Promotionsstudium als Stipendiat der International Max Planck Research School (IMPRS) „Uncertainty“ in Bonn und Jena; mehrere Auslandsaufenthalte v.a. in China; 2013–2015 Rechtsreferendariat beim Freistaat Thüringen; derzeit Referendar in Erfurt, Nachwuchskollegiat der Heidelberger Akademie der Wissenschaften und Gastforscher (Visiting Researcher) am Max-Planck-Institut zur Erforschung von Gemeinschaftsgütern in Bonn.

ISBN 978-3-16-153322-8 \ eISBN 978-3-16-159731-2 unveränderte Ebookausgabe 2020
ISSN 1614-8169 (Grundlagen der Rechtswissenschaft)

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliographie; detaillierte bibliographische Daten sind im Internet über *http://dnb.dnb.de* abrufbar.

Das Buch wurde von Gulde Druck in Tübingen auf alterungsbeständiges Werkdruckpapier gedruckt und gebunden.

Dem Andenken an
Theodore Eisenberg
(* 1947 † 2014)

Vorwort

Die vorliegende Arbeit ist zwischen Mai 2010 und März 2013 als Dissertation am Max-Planck-Institut zur Erforschung von Gemeinschaftsgütern in Bonn entstanden. Die Herausgeber der Schriftenreihe „Grundlagen der Rechtswissenschaft“ waren so freundlich, das Erscheinen der Arbeit in diesem honorigen Rahmen zu ermöglichen.

Finanziell erleichtert wurde die Veröffentlichung durch eine Publikationsbeihilfe der Deutschen Forschungsgemeinschaft (DFG), Bonn. Die Esche Schümann Commichau (ESC) Stiftung, Hamburg, hat die Arbeit im April 2014 durch ihren 88. Dissertationspreis gefördert. Daneben wurde die Arbeit im Mai 2014 mit dem Telekom-Preis für zivilrechtliche Dissertationen der Rechts- und Staatswissenschaftlichen Fakultät der Universität Bonn und im Juni 2014 mit der Otto-Hahn-Medaille für den wissenschaftlichen Nachwuchs der Max-Planck-Gesellschaft zur Förderung der Wissenschaften e.V. (MPG), München, ausgezeichnet. Von der Körber-Stiftung, Hamburg, wurde die Arbeit unter 438 eingereichten Dissertationen für das Finale des Deutschen Studienpreises im Juli 2014 ausgewählt.

Christoph Engel schrieb über wissenschaftliche Paradigmen einmal, sie seien „voraussetzungsvolle soziale Leistungen“. Das gilt für neue Paradigmen ebenso wie für jede andere wissenschaftliche Arbeit. Forschung setzt nicht nur einen sicheren Stand auf den Schultern des sprichwörtlichen Riesen voraus, sondern vor allem sozialen Austausch und die kontinuierliche Zusammenarbeit mit anderen. Die vorliegende Arbeit macht da keine Ausnahme. Zahlreichen Wegbegleitern ist zu danken, die mich unterstützten und auf deren Schultern die Arbeit ebenso ruht wie auf meinen.

In erster Linie danke ich den Mitarbeitern der Bibliothek im Bonner Max-Planck-Institut. Namentlich Julia Pagel, Sabrina van Detten, Regina Goldschmitt, Charlotte Mainzer, Paul Frühauf, Daniel Zimmel und Laura-Isabell Dietz standen mir allzeit hilfreich, zuverlässig und mit geradezu beängstigender Geschwindigkeit zur Seite, wann immer sich ein Krumen Weisheit in den Bibliotheken oder Bücherläden dieser Welt zu verstecken trachtete. Ebenso bei der Spurensuche halfen mir Sharon Wang und Sandra Geddes von der York University Law Library in Toronto und Elke Halsen-Raffel von der Bibliothek des Max-Planck-Instituts für ausländisches und internationales Privatrecht in Hamburg.

Daneben ist einer Vielzahl von Gleichgesinnten und Korrespondenzpartnern zu danken, die mir für konkrete Fragen zur Verfügung gestanden haben. Dazu gehören, in alphabetischer Reihung, Professor Carsten Burhop (Universität Wien), Professor Daylian Cain (Yale School of Management), Professor David Dunning (Cornell University), Professor Kent Greenfield (Boston College Law School), Professor Henry Hansmann (Yale Law School), Professor Heribert Hirte (Universität Hamburg), Professorin Poppy McLeod (Cornell University), Professor Alexander Pepper (London School of Economics and Political Science) und Professorin Christine Windbichler (Humboldt-Universität Berlin). Der Austausch mit ihnen allen machte mir erst deutlich, welche grenzüberschreitend herzliche Verbundenheit die globale Familie der Grundlagenforscher eint. Aus anderem Grund gebührt Professor Daniel Zimmer, Dr. Hendrik Wieduwilt und Rechtsanwalt Dr. Markus Englerth mein Dank: Sie haben unabhängig voneinander und ohne es zu wissen mein Interesse in die Richtung der vorliegenden Arbeit gelenkt. Dr. Gunnar Janson hat durch seine Arbeit zur „Ökonomischen Theorie im Recht“ meinen Untertitel inspiriert.

Wichtiger noch – ja absolut unentbehrlich – waren die vielen Gespräche, die ich mit Kollegen, Freunden und Förderern führen durfte. Hervorzuheben ist zunächst Professor Marcus Lutter, der auch die Zweitbegutachtung dieser Arbeit ohne Zögern übernommen und trotz sorgfältigster Lektüre beeindruckend schnell abgeschlossen hat. Hilfreiche Anregungen verdanke ich des Weiteren Ass. iur. Monia Manâa und Pascal Langenbach, die die Arbeit aufmerksam gegengelesen haben, sowie meinen weiteren Gesprächs- und Sparringspartnern Yoan Hermstrüwer, Dr. Kristoffel Grechenig, Dr. Alexander Morell, Dr. Sven Fischer und Dr. Sebastian Goerg vom Max-Planck-Institut Bonn, Dr. Klaus-Ulrich Schmolke und Dr. Eckart Bueren vom Max-Planck-Institut Hamburg sowie den Rechtsanwälten Dr. Nikolaus Reinhuber und Dr. Manuel Lorenz in Frankfurt sowie Dr. Arnold Mock in Düsseldorf. Einen besonders wichtigen Impuls für die thematische Fokussierung der Arbeit verdanke ich Dr. med. Konrad Stopsack.

Der größte Dank aber gebührt meinem Betreuer Professor Christoph Engel für den Wagemut, einen in empirischen Fragen völlig unbedarften Jungjuristen in ein Institut aufzunehmen, das Heteroskedastizität nicht nur lehrt, sondern lebt. „Sie werden das experimentelle Handwerkszeug erst lernen müssen“, schrieb er mir nach dem Bewerbungsgespräch, „aber in Ihrem Vortrag und beinahe mehr noch in unserem Gespräch im Anschluss haben Sie mir den Eindruck vermittelt, dass das für Sie keine Pflichtübung wäre, sondern dass Sie das auch wirklich wollen.“ Christoph Engel hat mich damals richtig eingeschätzt und seitdem meine akademischen Gehversuche in jeder erdenklichen Weise gefördert – ideell wie finanziell. Dafür sowie für seine stete Bereitschaft zum Gespräch und zur Vorablektüre kann ich ihm nicht genug danken.

Zum Schluss bleibt nur noch denen zu danken, denen man nicht danken *kann*, jedenfalls nicht angemessen: Dr. Christiane Heinicke für Musen- und Mußestunden und das Erdulden von weit mehr als einer unruhigen Nacht, die allein der Titel dieser Arbeit gekostet hat, sowie sechs weiteren Menschen, die mich erst auf den langen Weg gebracht haben, von dessen Etappenziel ich nun zurückblicke: Kornelia und Dr. Ingo Hamann, Don Vinge, Dorothea Hoffmann, Professor Heribert Heckschen und Rechtsanwalt Dr. Hans-Christoph Ihrig. Sie wissen wofür.

Erfurt, im August 2014 *Hanjo Hamann*

Inhaltsübersicht

Inhaltsverzeichnis

Hinweise zur Quellenarbeit

„Die Dissertation ist in deutscher Sprache abzufassen", heißt es in § 6 S. 1 der Juristischen Promotionsordnung der Universität Bonn. Dem will die vorliegende Arbeit auch in ihrem Umgang mit Primärquellen gerecht werden. Im Haupttext taucht die englische Sprache daher allenfalls in Klammerzusätzen auf, im Übrigen habe ich Fachbegriffe, die noch keine deutsche Entsprechung hatten, ins Deutsche übertragen und allfällige Zitate übersetzt, wo die ursprüngliche Quelle nicht in einer übersetzten Ausgabe zitiert werden konnte.

Dennoch verlangt die vorliegende Arbeit nach dem Originalton der verwendeten Quellen, und zwar sowohl im Interesse der Authentizität als auch im Interesse des deutschen Lesers, dem nicht alle ausländischen Quellen gleichermaßen zugänglich sein werden. Daher habe ich die Fußnoten mehr als üblich mit wörtlichen Zitaten in der Originalsprache gespickt und die Zitierweise (Hervorhebungen usw.) unverändert belassen, soweit nicht anders angemerkt. Auslassungen habe ich mit […] kenntlich gemacht, bei genau einem Wort mit dem weniger gebräuchlichen Auslassungszeichen [.].

Um trotz der Vielfalt der verwendeten Quellen ein einheitliches und unmissverständliches Zitierformat zu verwenden, das jede Quelle möglichst ohne Konsultation des Literaturverzeichnisses direkt aus der Fußnote auffindbar macht, bin ich wie folgt verfahren:

Um Buch- von Zeitschriftenveröffentlichungen unterscheiden zu können, habe ich Buchveröffentlichungen mit einem bis drei vollständig ausgeschriebenen Titelworten sowie dem Erscheinungsjahr der neuesten Ausgabe zitiert und Abkürzungen bewusst unterlassen, mit Ausnahme amtlicher Gesetzeskürzel[1] sowie der sechs Abkürzungen AG (Aktiengesellschaft), AktR (Aktienrecht), FS (Festschrift), GftsR (Gesellschaftsrecht), Hdb (Handbuch) und MK (Münchener Kommentar).

Zeitschriften habe ich durchweg einheitlich zitiert unter Nennung von Autor(in), Abkürzung des Zeitschriftentitels, Jahrgang und Anfangs- sowie Zitatseite. Dabei habe ich deutsche Zeitschriftentitel so abgekürzt wie in Deutschland gebräuchlich und aus einschlägigen Verzeichnissen ersicht-

---

[1] Online nachgewiesen in *BMJ*, Gesetze 2012.

lich.[2] Dasselbe gilt für einzelne englischsprachige aber aus Deutschland betreute Zeitschriften (insbesondere ECFR,[3] JITE,[4] EBOR,[5] und sbr[6]). Ausländische Zeitschriftentitel dagegen habe ich in einem eigenen Format abgekürzt, das sich lose an internationale Zitierweisen anlehnt,[7] aber viele Abkürzungen auflöst, die in einzelnen Fachdisziplinen etabliert sind (z.B. AER und JPSP)[8].

Dabei kamen vor allem folgende Kürzel zum Einsatz:

Formate
Ann.............Annals
Bull.............Bulletin
J..................Journal
Q.................Quarterly
Rev.............Review

Disziplinen
Acc.............Accounting
Biol.............Biology / ~ical
Bus.............Business
Econ...........Economic(s)
Edu.............Education(al)
Fin..............Finance / ~ial
Hist.............History / ~ical
L.................Law / Legal
Med.............Medicine
Mgmt..........Management
Polit............Political
Psy..............Psychology / ~ical
Sociol..........Sociology / ~ical
Stat.............Statistics / ~ical
Tech............Technology

Regionen
Am.............American
Brit.............British
Can.............Canadian
Eur.............European
Ger.............German
Int...............International

Methoden
Appl............Applied
Clin.............Clinical
Comp..........Comparative
Emp............Empirical
Eval.............Evaluation
Exp.............Experimental
Gen.............General
Lit...............Literature
Meas...........Measurement
Meth...........Methods / ~ology
Sci...............Science
Stud.............Studies
Res..............Research
Surv............Survey

---

[2] *Kirchner/Pannier*, Abkürzungsverzeichnis 2012; *juris GmbH*, Verzeichnis 2014; *Kuselit Verlag*, Gesamtliste 2014.

[3] *European Company and Financial Law Review*, engl. Äquivalent zur Zeitschrift für Unternehmens- und Gesellschaftsrecht (ZGR).

[4] *The Journal of Institutional and Theoretical Economics*, bis 1986 Zeitschrift für die gesamte Staatswissenschaft (ZgS).

[5] *European Business Organization Law Review.*

[6] *Schmalenbach Business Review*, engl. Pendant zur Zeitschrift für betriebswirtschaftliche Forschung (zfbf).

[7] Vor allem *Prince*, Bluebook 2010, 444 ff. (Tab. 13) und *Thomson Reuters*, Abbreviations 2012.

[8] *American Economic Review* (hier als Am Econ Rev) und *Journal of Personality and Social Psychology* (hier als J Pers Soc Psy) gehören zu den renommiertesten Zeitschriften der Ökonomik bzw. Sozialpsychologie.

Schlagworte

Assoc	Association	Manage	Managerial
Admin	Administrative	Mktg	Marketing
Adv	Advances	Org	Organizational
Ann	Annual	Pers	Personality
Behav	Behavior(al)	Persp	Perspectives
Cog	Cognitive	Proc	Processes
Comm	Communication	Publ	Public
Corp	Corporate / ~ion	Rel	Relations
Dec	Decision	Soc	Society / Social
Hum	Human	Strat	Strategic
Inf	Information	U	University

Nach *Stiglers Gesetz* trägt keine wissenschaftliche Entdeckung den Namen ihres wahren Entdeckers.[9] Wo also Konzepte den Namen einer historischen Persönlichkeit tragen, die nach aktuellem Erkenntnisstand nicht als erster Entdecker bzw. Erfinder angesehen werden kann, habe ich versucht, dem Ersterfinder dadurch gerecht zu werden, dass ich seinen Namen der herkömmlichen Bezeichnung vorangestellt habe, etwa im Fall der de Moivre-Gaußschen Glockenkurve (§ 2 Fn. 110), des Kaplan-Masowschen Hammers (§ 2 Fn. 250) oder des Yule-Simpson-Paradox (§ 3 bei Fn. 83).

Alle Quellen sind auf dem Stand der Fertigstellung dieser Arbeit (Anfang 2013); obwohl zahlreiche Aktualisierungen nachgetragen wurden, konnte eine systematische Nachrecherche nicht erfolgen.

---

[9] *Stigler*, Trans NY Acad Sci 1980, 147: "No scientific discovery is named after its original discoverer." – Stigler versteht sein Gesetz als "self-proving theorem", denn es stamme seinerseits von Robert Merton: "If there is an idea in this paper that is not at least implicit in Merton's *The Sociology of Science*, it is either happy accident or a likely error." (ebd. 147); weitere Bsp. bei *Salsburg*, Statistics 2001, 16 Fn. 1, 26 Fn. 1, 30 Fn. 2, 33 Fn. 1, 68 Fn. 1, 77 Fn. 1, 112 Fn. 2 usw.

§ 1

Standortbestimmung und Untersuchungsprogramm

„Wie viel besser wäre die Rechtswissenschaft daran, könnte sie, wie die Naturwissenschaften, unmittelbar an den Gegenstand herantreten.“[1]
– Julius von Kirchmann 1848

„Rechtswissenschaft muss daher zugleich *Seinswissenschaft* und *Normwissenschaft* sein. Sie ist in sich bereits interdisziplinär.“[2]
– Philippe Mastronardi 2003

Wenn Recht nicht nur gerecht, sondern auch *sach*gerecht sein will, kommt es nicht umhin, *Sach*lagen zu erfassen, also Erfahrungswissen zu berücksichtigen. Welche Methoden stehen dafür zur Verfügung, und wo liegen ihre Grenzen? Das ist die Frage, die hier zu beantworten ist.

Dass Juristen sich mit Erfahrungswissen befassen, ist nichts grundlegend Neues. „Die Rechtswissenschaft war schon immer eine neugierige Wissenschaft“,[3] und auch „schon immer eine anlehnungsbedürftige Disziplin.“[4] Viele helle Köpfe haben in verdienstvollen Arbeiten versucht, den Austausch der Rechtswissenschaft mit Erfahrungswissenschaften zu vermitteln, ohne einander in die „ausschließliche Unterwerfung unter fremdwissenschaftliche Ergebnisse“[5] zu drängen. Neu ist dagegen das umfassende Programm einer methodisch reflektierten *evidenzbasierten Jurisprudenz*. Die Verbindung der juristischen Lebenserfahrung mit den Erkenntnissen der empirischen Forschung kann die Sachgerechtigkeit juristischer Entscheidungen fördern und die juristische Forschung und Praxis unterstützen. Dazu bietet die vorliegende Arbeit eine praktische Handreichung und kritische Reflexion.

---

[1] *von Kirchmann*, Jurisprudenz 1848, 16, aber 17: „beinahe überall im Recht hat das Gefühl sich schon für eine Antwort entschieden, ehe noch die wissenschaftliche Untersuchung begonnen hat.“

[2] *Mastronardi*, Juristisches Denken 2003, Rn. 287 und weiter Rn. 720: „Juristisches Denken wird seiner interdisziplinären Aufgabe nur gerecht, wenn es Norm und Realität, Logik und Empirie zugleich umfasst.“

[3] *Lüdemann* in: Engel u.a., Recht und Verhalten 2007, 1.

[4] *Engel/Schön* in: Engel/Schön, Proprium 2007, IX.

[5] *Schön* in: Engel/Schön, Proprium 2007, 313, 316.

A. Evidenzbasierte Jurisprudenz

Der englische Volksmund warnt davor, Bücher nach ihrem Titel zu beurteilen (*Don't judge a book by its cover*). Das ist hier ohnehin schwierig, denn das erste Wort im Titel dieses Buches ist keine fünfundzwanzig Jahre alt und im Deutschen noch nie im Zusammenhang mit der Jurisprudenz oder mit Recht oder Rechtswissenschaft aufgetaucht.[6] Im englischsprachigen Schrifttum werden zwar seit knapp zehn Jahren sporadisch Zusammenhänge zwischen Recht, Rechtsetzung und Evidenzbasierung aufgezeigt, aber nicht im hier angedachten Sinn.[7] Auch die Zusammensetzung „evidenzbasierte Jurisprudenz" war bis zur Fertigstellung dieser Arbeit im englischen Sprachraum nicht geläufig.[8] Daher ist zunächst der *Begriff* „Evidenzbasierung" zu erläutern (I.), bevor das *Programm* der evidenzbasierten Jurisprudenz vorgestellt (II.) und der weitere Gang der Darstellung (III.) entworfen werden kann.

I. Der Begriff

Der Begriff der „Evidenz" (von lat. *ex* und *videns*, heraus sehend)[9] ist in unglücklicher Weise mehrdeutig. Unglücklich deshalb, weil seine verschiedenen Bedeutungen einander diametral widersprechen: Was vorliegend als „Evidenz" bezeichnet wird, ist das genaue Gegenteil dessen, was Juristen gewöhnlich darunter verstehen. Bevor also der hier verwendete Evidenzbegriff erläutert wird, auf den sich das Adjektiv „evidenzbasiert" bezieht, wird zur begrifflichen Abgrenzung zunächst das juristische Begriffsverständnis skizziert; offenbleiben muss die Frage, welches Begriffsverständ-

---

[6] Für die Zusammensetzung jedes dieser drei Substantive mit dem Adjektiv „evidenzbasiert" verzeichnen Google, Google Scholar, juris, Beck Online und Kuselit Online jeweils null Treffer, Stand 12.12.2012.

[7] *Trujillo*, UCLA L Rev 2005, 357, 382 Fn. 80 entwickelte einen "evidence-based law approach", allerdings "by treating doctrine as a quantitative unit" (363); *Van Gestel* in: Schäffer/Iliopoulos-Strangas, SIPE 2007, 139, 142 versteht unter "evidence-based lawmaking" die Rechtssetzung "on a trial and error basis" – ebenso *Davis*, U Toronto L J 2010, 537, 548 ("legal experimentalism"); *Cook/Dickens/Erdman*, Int J Gyn Obstet 2006, 191, 192 f. und ausf. *Rachlinski*, Cornell L Rev 2011, 901 erörterten "evidence-based law" als *Produkt* der empirischen Rechtsforschung; ähnl. *Posner*, Duq L Rev 2013, 316: "We need evidence-based law, just as we need evidence-based medicine."; näher am hiesigen Verständnis eines *Verfahrens Posner*, Geo L J 2009, 845, 852: "placing law on a solid empirical basis (the analogy is to the movement for 'evidence-based' medicine)".

[8] Sie tauchte zwar am Rande im Leitartikel *o.V.*, Can Med Assoc J 1999, 229 auf (im Sinne folgenorientierter Rechtsprechung), bewusst reflektiert aber erst *Girvan/Deason*, Clev St L Rev 2013, 1057 zum "acknowledgement that evidence-based jurisprudence is desirable".

[9] *Harper*, Etymology 2014, Stichworte "evidence" und "evident".

nis historisch vorgängig war, oder ob sie sich unabhängig voneinander aus der selben lateinischen Begriffswurzel entwickelt haben.[10]

Der „Rechtsbegriff der Evidenz“ dürfte gut ein halbes Jahrhundert alt sein und stützt sich auf philosophische Vorarbeiten.[11] Er bedeutet so viel wie „Offensichtlichkeit“ und bezeichnet Tatbestände, die „unmittelbar einleuchtend“ und „so augenscheinlich“ sind, dass es „keiner weiteren Schlüsse bedarf“, um sie „Gewißheit werden zu lassen“.[12] „Evidenz“ im juristischen Sinn des Wortes ist also „in eminenter Weise von Erfahrung und Sachgefühl abhängig“[13] und beruht maßgeblich auf Introspektion:

> „Wer die Evidenz bemüht, beruft sich auf die eigene Einsicht […] der es am Element der intersubjektiven Nachvollziehbarkeit mangelt. […] Als ‚evident‘ bezeichnen wir gewöhnlich Behauptungen, die wir einfach unserem […] Wissen entnehmen, ohne daß diese Entnahme eine thematische Reflexion voraussetzte.“[14]

Wenngleich in diesem Zusammenhang auch von einer „Evidenzbasis“ die Rede ist,[15] hat das mit dem hier verwendeten Adjektiv „evidenzbasiert“ nichts zu tun.[16] Denn bei der „Evidenz“ im juristischen Sprachverständnis geht es „nie um die Evidenz von Fakten, sondern immer um die Evidenz von Rechtlichem.“[17] Die vorliegende Arbeit dagegen bezeichnet mit Evidenz *gerade* Wissen über tatsächliche Zusammenhänge, und auch nur soweit es durch systematische Verfahren zur „Veranschaulichung“[18] oder zum „Nachweis“,[19] und nicht allein durch Introspektion, gewonnen wird.

Die Wortzusammensetzung „evidenzbasiert“ (*evidence-based*) kommt aus der Medizin. Dort bezeichnet sie die „gewissenhafte, ausdrückliche

---

[10] Laut *Harper*, Etymology 2014, Stichwort “evidence”, ist die Bedeutung „Nachweis“ früher belegt (im späten 14. Jh.) als „Offensichtlichkeit“ (um 1660).

[11] *Achterberg*, DÖV 1963, 331.

[12] Ausf. *Achterberg*, DÖV 1963, 331, 332; *Scheuerle*, ZZP 1971, 241; *Mayer-Maly* in: Marcic u.a., FS Verdroß 1971, 259; *Schreiner* in: Krawietz u.a., GS Tammelo 1984, 543; *Krugmann*, Evidenzfunktionen 1996; vgl. auch *Creifelds/Weber*, Rechtswörterbuch 2007, 390 („Evidenzprüfung“); *Köbler*, Etymologisches Wb. 1995, 118; *Müller/Christensen*, Methodik I 2013, 269 ff. (Rn. 258–261); *Welti/Raspe*, NJW 2002, 874 bei Fn. 11 und 12.

[13] *Mayer-Maly* in: Marcic u.a., FS Verdroß 1971, 259, 263, 265, 266 f.

[14] *Schreiner* in: Krawietz u.a., GS Tammelo 1984, 543 sowie 549.

[15] *Schreiner* in: Krawietz u.a., GS Tammelo 1984, 543, 544, 546.

[16] Zugleich widerspricht es dem hiesigen Begriffsverständnis nicht zwingend, denn *Schreiner* in: Krawietz u.a., GS Tammelo 1984, 543, 546 meint mit „Evidenzbasis“ jedes „Fundament von Gewißheit“ – auch eines, das auf „Erfahrung“, „Beobachtung“ oder „Sinnesdaten“ beruht.

[17] *Mayer-Maly* in: Marcic u.a., FS Verdroß 1971, 259, 264, auf 263 näher zu den drei „Schwerpunkten des Gebrauchs von evidens“: Willensinhalten, Rechtssätzen und Werturteilen.

[18] *Dudenredaktion*, Fremdwörterbuch 2000, 421 („Evidenz“).

[19] *Stein*, DÄBl 1998, A 862, 864; *Bilger* in: Comberg/Klimm, Allgemeinmedizin 2004, 74; ähnl. Bundesanzeiger-Verlag, www.bundesanzeiger-verlag.de/service/evidenzzentrale.html.

und umsichtige Verwendung der aktuell besten Nachweise“ für die Behandlung von Patienten[20] und wurde dem Fachpublikum 1992 durch das neuartige Unterrichtskonzept einer kanadischen Universität bekannt.[21]

Danach verbreitete sich die Wortschöpfung rasant in der klinischen Praxis,[22] fand ihren Weg in die Gesundheitswirtschaft[23] und hielt in vielen sozialplanerischen Anwendungsdisziplinen Einzug – etwa der Kriminalprävention,[24] der Bildungsforschung,[25] der allgemeinen Politikfeldforschung,[26] sowie der Betriebs-[27] und Volkswirtschaftslehre.[28] Dieser rasante Siegeszug der Evidenzbasierung und die bemerkenswerten Blüten, die sie am Rande ihres Weges treibt („evidenzbasiertes Bibliothekswesen“,[29] „evidenzbasierte Kochkunst“,[30] etc.) belegen, dass Evidenzbasierung längst zum aktuellen Zeitgeist gehört:

---

[20] *Sackett/Rosenberg u.a.*, Brit Med J 1996, 71: “Evidence based medicine is the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients.”; zit. von *Bilger* in: Comberg/Klimm, Allgemeinmedizin 2004, 74; *Windeler*, Gesundheitswesen 2008, 418; *Stallberg*, PharmR 2010, 5, 6; identisch *Welti/Raspe*, NJW 2002, 874.

[21] *EBM Working Group*, J Am Med Assoc 1992, 2420; erstmals verwendet wurde der Begriff “evidence-based medicine” schon 1990/91, so ebd. 2421 Fn. 18 sowie *Jaeschke/Guyatt*, Seminars Med Pract 1999, H. 3, 3 m.w.N.; auf historische Vorläufer verweisen *Raspe*, ZaeF 1996, 553, 555; *Rangachari*, J Royal Soc Med 1997, 280 f.

[22] *Rosenberg/Donald*, Brit Med J 1995, 1122; *Sackett/Rosenberg u.a.*, Brit Med J 1996, 71 – teilweise übersetzt in *Raspe*, ZaeF 1996, 553; *Stein*, DÄBl 1998, A 862; *Straus/Jones*, Brit Med J 2004, 987 und die vier nachfolgenden Leitartikel im selben Themenheft; *Sawicki*, DÄBl 2005, A 888; *Raspe*, GesR 2012, 584.

[23] *Hewison*, J Nurs Mgmt 1997, 195; *Kovner/Elton/Billings*, Front Health Serv Mgmt 2000, 3 (mit 4 Repliken ab S. 25 und Duplik auf S. 45 f.); *Walshe/Rundall*, Milbank Q 2001, 429, 437 m.w.N.; *Young*, J Nurs Mgmt 2002, 145.

[24] *Sherman*, Policing 1998; *MacKenzie*, Crime Delinq 2000, 457; *Sherman* in: Waring/Weisburd, Crime 2002, 217; *Sherman*, Prevention 2002; *Bilsky/Döring u.a.*, FPPK 2012, 69.

[25] *Slavin*, Edu Res 2002, 15; *Davies*, Brit J Edu Stud 1999, 108; *Bridges/Smeyers/Smith*, Education 2009; *Spiel*, PsR 2009, 255; *Otto/Polutta/Ziegler*, Soziale Arbeit 2010; spezifisch juristisch *Zimmerman*, Seattle U L Rev 2012, 305, 366; *Winteler/Forster* in: Brockmann u.a., Methoden 2012, 20.

[26] *Ham/Hunter/Robinson*, Brit Med J 1995, 71; *Black*, Brit Med J 2001, 275; *Davies/Nutley/Smith*, What Works? 2000; *Sanderson*, Publ Admin 2002, 1; *Rüb/Straßheim* in: Geis u.a., Legitimitätspolitik 2012, 377; *Dicks*, Nat 2013, 283.

[27] *Pfeffer/Sutton*, Harv Bus Rev 2006, 62; *Pfeffer/Sutton*, Hard Facts 2006; *Rousseau*, Acad Mgmt Rev 2006, 256; ebenso zahlreiche Veröffentlichungen seit 2006 von *J. Scott Armstrong*: https://marketing.wharton.upenn.edu/profile/226/research.

[28] Seit Herbst 2013 bietet die Universität München einen Studiengang “Evidence-Based Economics” an, vgl. www.evidence-based-economics.de.

[29] *Lerdal*, L Lib J 2006, 33.

[30] *Enserink*, Sci 2006, 1235.

„Die Phrase ‚evidenzbasiert' ist ein Modewort der aktuellen politischen Debatte und steht unter dem Damoklesschwert all der Banalität und Oberflächlichkeit, die mit Modewörtern einhergehen. Lassen wir uns nicht von ihrer aktuellen Beliebtheit irreführen: Hinter evidenzbasierter Praxis stehen vor allem gehörige Substanz und Disziplin."[31]

Das Zusammenspiel von Substanz und Disziplin im Verfahren der Evidenzbasierung wird besonders an der Medizin deutlich: Obwohl sie „alle anderen Berufsstände darin übertrifft, massenweise hochwertige Nachweise zu produzieren" und ihre empirische Substanz schon 1995 über eine Million veröffentlichter klinischer Versuche umfasste, hielt deren Verwendung in der ärztlichen Entscheidungsfindung „nicht einmal annähernd mit".[32] Dementsprechend stellte evidenzbasierte Medizin (EbM) ursprünglich den disziplinierenden Versuch dar, Intuition und „unsystematische klinische Erfahrung" in der ärztlichen Entscheidungsfindung zurückzuschneiden und durch „Nachweise aus der klinischen Forschung" zu ersetzen.[33] An die Stelle des mit der Autorität der Erfahrung vorgetragenen ärztlichen Expertenwissens („oder ähnlichen Diskussionsstoppern"[34]) sollte also die mit der Autorität der Methodenkompetenz vorgetragene wissenschaftliche Erkenntnis treten.[35] Dieses neue Verständnis von der angemessenen Tatsachengrundlage klinischer Entscheidungen verbuchte schnell Erfolge:

„Die stärksten Argumente für eine evidenzbasierte Medizin kommen von den vielen Beispielen, bei denen sich als sicher oder nützlich geltende Interventionen nach entsprechender Evaluation als schädlich oder unwirksam erwiesen haben."[36]

---

[31] *Rousseau*, Acad Mgmt Rev 2006, 256, 258; die allgemeine Forderung nach einer "evidence-based society" erhob schon *Smith*, J Royal Stat Soc A 1996, 367.

[32] *Sherman* in: Waring/Weisburd, Crime 2002, 217, 222 f.; zynisch *Pfeffer/Sutton*, Harv Bus Rev 2006, 62, 2: "If all this sounds laughable to you – after all, what else besides evidence *would* guide medical decisions? – then you are woefully naive about how doctors have traditionally plied their trade. […] here's what doctors rely on [.]: obsolete knowledge gained in school, long-standing but never proven traditions, patterns gleaned from experience, the methods they believe in and are most skilled in applying, and information from hordes of vendors with products and services to sell."

[33] *EBM Working Group*, J Am Med Assoc 1992, 2420; ebenso *Guyatt/Cook/Haynes*, Brit Med J 2004, 990: "In its original formulation, this discipline reduced the emphasis on unsystematic clinical experience and pathophysiological rationale, and promoted the examination of evidence from clinical research."

[34] *Augsberg*, GesR 2012, 595, 600.

[35] Krit. aber *Feinstein/Horwitz*, Am J Med 1997, 529, 534 zur mangelnden Transparenz der "new group of experts".

[36] *Timmer*, DÄBl 2012, A 1418, 1420; *Windeler*, Gesundheitswesen 2008, 418, 426: „Neue Ideen sind oft eine Reaktion auf alte Missstände. Es sagt viel über die Medizin als solche, dass sie eines Attributs wie evidenzbasiert überhaupt bedarf."

Daher wurde Evidenzbasierung vollmundig als Paradigmenwechsel ausgerufen,[37] und weckte – wie jeder Herausforderer herrschender Paradigmen – zahlreiche Befürchtungen, die vom Einzug bürokratischer Kochbuchmentalität über die Entmündigung praktizierender Ärzte bis hin zur wirtschaftlichen Instrumentalisierung der Forschung reichten.[38] Diese Kritik hat dem Siegeszug der Evidenzbasierung keinen Abbruch getan, aber einige ernstzunehmende Schwächen aufgedeckt und so zu ihrer Fortentwicklung beigetragen.[39] Heute wird Evidenzbasierung nicht länger als radikale Alternative zur traditionellen Medizin verstanden,[40] sondern als integrierender Ansatz, der auf zwei Säulen ruht: Ärztlicher Erfahrung *und* klinischen Studien, die auch als „innere" und „äußere Expertise"[41] oder „interne und externe Evidenz" bezeichnet werden.[42] Die evidenzbasierte ist also Ergänzung, nicht Ersatz, für die klassisch „eminenzbasierte"[43] Medizin.

Evidenzbasierung ist daher nicht mehr und nicht weniger als die Suche nach sachangemessenen Quellen der Wirklichkeitserkenntnis durch bewusste Reflexion ihrer Möglichkeiten und Grenzen.[44] Dazu werden verschiedene Erkenntnisquellen nach ihrer allgemeinen „Überzeugungskraft"[45] (oder umgekehrt: „Fehleranfälligkeit"[46]) beurteilt und mehrere Evidenzstufen gebildet (*levels of evidence*), die einerseits eine Hierarchisierung begründen, andererseits aber als komplementäre Schritte (*steps*) bei

---

[37] *EBM Working Group*, J Am Med Assoc 1992, 2420 m.w.N. in Fn. 4 zu der von *Thomas Kuhn* begründeten Theorie des wissenschaftlichen Fortschritts, aus der das Konzept des Paradigmenwechsels stammt.

[38] *Raspe*, ZaeF 1996, 553, 8 f.; ausf. *Feinstein/Horwitz*, Am J Med 1997, 529, passim sowie *Straus/McAlister*, Can Med Assoc J 2000, 837 (Literaturbericht über 47 EbM-kritische Aufsätze); dagegen *Sackett/Rosenberg u.a.*, Brit Med J 1996, 71, 72; *Reilly*, Brit Med J 2004, 991, 992; *Bilger* in: Comberg/Klimm, Allgemeinmedizin 2004, 74.

[39] *Guyatt/Cook/Haynes*, Brit Med J 2004, 990: "The philosophy of evidence based medicine has evolved. Exponents increasingly emphasise the limitations of using evidence alone to make decisions".

[40] *Rangachari*, J Royal Soc Med 1997, 280, 283: "the designation of one system of medicine as evidence-based implies that other systems are not based on evidence. This is far from true."

[41] *Bilger* in: Comberg/Klimm, Allgemeinmedizin 2004, 74.

[42] *Stallberg*, PharmR 2010, 5, 6 m. Verw. auf *Sackett/Rosenberg u.a.*, Brit Med J 1996, 71: "The practice of evidence based medicine means integrating individual clinical expertise with the best available external clinical evidence from systematic research."; ebenso *Windeler*, Gesundheitswesen 2008, 418, 420 f.; ähnl. *Greenhalgh*, Brit Med J 1999, 323.

[43] *Windeler*, Gesundheitswesen 2008, 418.

[44] *Rangachari*, J Royal Soc Med 1997, 280, 283: "What is at stake is not the presence or absence of evidence but the legitimacy of the evidence used to justify [.] decisions."

[45] *Welti/Raspe*, NJW 2002, 874.

[46] *Stallberg*, PharmR 2010, 5, 8.

der Entscheidungsfindung einander gerade nicht ausschließen sollen.[47] Entscheidend ist die bewusste und „ausdrückliche“ Reflexion jedes einzelnen Erkenntnisschrittes.[48] Dieses „zentrale Prozeduralisierungselement macht die EBM für Juristen verständlich und anschlussfähig.“[49]

II. Das Programm

Mit dem deutschen Recht kam die Evidenzbasierung erstmals durch das Gesetz zur Reform der gesetzlichen Krankenversicherung 2000 in Berührung, das weltweit wohl zum ersten Mal überhaupt[50] die Evidenzbasierung als gesetzlichen Maßstab festschrieb (§ 137e III 1 Nr. 1 SGB V i.d.F. 2000–2003). Was zuvor noch „als rein intraprofessionelles Unternehmen, als Selbstvergewisserung und Selbstbindung der klinischen Professionen erschien“,[51] war nun plötzlich als Regulierungskonzept salonfähig geworden.[52] Zwar wurde „evidenzbasiert“ anfangs gleichgesetzt mit „wissenschaftlich fundiert“,[53] was den eben dargestellten Zwei-Säulen-Ansatz zu verkürzen droht. Gleichwohl findet sich die Evidenzbasierung mittlerweile an zahlreichen Stellen des Fünften Sozialgesetzbuches (und anderer Gesetze[54]) und wurde von der Rechtsprechung grundsätzlich anerkannt.[55] Freilich bleibt auch von der Herausbildung eines „einheitlichen Rechtsbegriffs der Evidenzbasierten Medizin“[56] noch ein großer Schritt zur Evidenzbasierung in der Jurisprudenz.

Dieser Schritt beginnt mit der Erkenntnis, dass Jurisprudenz und Medizin durch ein gemeinsames Erkenntnisinteresse verbunden sind:

---

[47] *Howick/Chalmers u.a.*, OCEBM Levels 2011; vgl. auch *Windeler*, Gesundheitswesen 2008, 418, 423 f. zu alternativen Klassifikationen (“SIGN” und “GRADE”).

[48] *Windeler*, Gesundheitswesen 2008, 418, 419 m.Verw. auf *Sackett/Rosenberg u.a.*, Brit Med J 1996, 71.

[49] *Augsberg*, GesR 2012, 595, 600.

[50] *DNEbM*, Chronik 2013.

[51] *Raspe*, GesR 2012, 584, 589, und weiter: „die resultierende Evidenz [blieb] sozusagen in der Familie.“

[52] *Timmer*, DÄBl 2012, A 1418, 1422: „EbM ist da. In der Sozialgesetzgebung, im Gemeinsamen Bundesausschuss, in der Qualitäts- und Wirtschaftlichkeitsbewertung im Gesundheitswesen (Stichwort: IQWiG), in den Empfehlungen der Fachgesellschaften (Stichwort: Leitlinien). Das ‚System‘ hat es weitaus schneller umarmt, als der am Krankenbett Tätige, für den EbM ursprünglich gedacht war.“

[53] Begr. § 137e SGB V-E, BuB Ausschuss für Gesundheit, BT-Drs. 14/1977 v. 3.11.1999, 171; laut *Timmer*, DÄBl 2012, A 1418 bevorzugen mittlerweile auch Mediziner den Begriff „wissenschaftlich begründete Medizin“.

[54] Vgl. *Schmidt-Recla* in: Rauscher, MK-FamFG 2013, § 280 Rn. 22 zu § 280 III Nr. 2 FamFG: „Dem Gesetzgeber schwebte mit dem Abstellen auf die ‚zugrunde gelegten Forschungserkenntnisse‘ wohl so etwas wie Evidenzbasierung vor.“

[55] BVerfGE 115, 25, 46 f.; dazu und zur Rspr. des BSG *Augsberg*, GesR 2012, 595, 597 ff.

[56] *Stallberg*, PharmR 2010, 5, 9.

„Das Fach heißt Jurisprudenz, nicht Jurisszienz. Die Juristerei ist also ein praktisches Fach […] steht offensichtlich also in einer Nähe zu den beiden anderen großen praktischen Disziplinen, der Technik und der Medizin.“[57]

Denn Juristen wie Ärzte interessieren sich für die Wirklichkeit (nur) insoweit, wie es für ihre praktischen Entscheidungen nötig ist.[58] Dabei üben und praktizieren Juristen wie Ärzte traditionell an pathologischen Fällen:[59]

„Der Patient erwartet für seine Heilbehandlung eine eingehende Diagnose, eine Analyse aller Daten und eine individuelle, auf ihn zugeschnittene Therapie. Ebenso verhält es sich mit der Rechtsanwendung. […] ein ‚makelloser‘, gewissermaßen ‚gesunder‘ Fall würde nicht zum Gegenstand eines […] Verfahrens werden.“[60]

Außerdem können Juristen wie Ärzte ihre Entscheidungen nicht vertagen, bis sie die Wirklichkeit vollständig begreifen.[61] Folglich liegt die Vermutung nahe, dass Juristen und Ärzte in ähnlicher Weise mit der Wirklichkeit umgehen müssen:

„Es wäre eines eigenen Projekts wert, die Beweismittel, Beweisermittlungs- und Beweiswürdigungsverfahren der Handlungswissenschaft Medizin und der (Sozial-)Rechtsprechung systematisch miteinander zu vergleichen.“[62]

Solange dieser Vergleich aussteht,[63] geben die aufgezeigten Parallelen immerhin Anlass, über die mögliche Rolle der Evidenzbasierung in der Jurisprudenz nachzudenken.

Um ein mögliches Missverständnis gleich vorweg auszuräumen: Evidenzbasierung hat nichts mit der Tatsachenermittlung in konkreten Einzelfällen zu tun. Das Augenmerk der Evidenzbasierung liegt nicht auf „singularen Tatsachenurteilen über den ‚Tatbestand‘“, sondern auf „generellen Tatsachenurteilen“,[64] also „über die Tatsachen des Einzelfalls hinausgrei-

---

[57] *Engel/Schön* in: Engel/Schön, Proprium 2007, IX, XII; ähnl. *Wissenschaftsrat*, Rechtswissenschaft 2012, 5: „Die Rechtswissenschaft gehört wie […] die Medizin zu den so genannten Professionsfakultäten. Als akademische Fächer kennzeichnet sie eine enge Theorie-Praxis-Verklammerung“; *Raspe*, GesR 2011, 449 fasst sie unter „humane Praxiswissenschaften“.

[58] Vgl. *Raspe*, GesR 2012, 584 und *Welti/Raspe*, NJW 2002, 874: „Handlungswissenschaften (wie klinische Medizin, Pädagogik und Jurisprudenz), in denen unter Unsicherheit auf der Basis von Wahrscheinlichkeitswissen entschieden werden muss.“

[59] *Engel/Schön* in: Engel/Schön, Proprium 2007, IX, XII; *Augsberg*, GesR 2012, 595, 599 zur „Anomalitätsorientierung“ der Medizin.

[60] *Jochum*, Grundfragen 2012, 152 f.

[61] *Augsberg*, GesR 2012, 595, 599; *Raspe*, GesR 2011, 449; *Raspe*, GesR 2012, 584: „Ihre Agenten müssen handeln“; vgl. auch *Feinstein/Horwitz*, Am J Med 1997, 529, 533 für die Medizin und unten Fn. 187 für die Rechtswissenschaft.

[62] *Raspe*, GesR 2011, 449, 451.

[63] Allg. aber *Upmeier*, Fakten 2010, v.a. 147 ff. über „juristische und wissenschaftliche Tatsachenfindung“.

[64] *Winter*, Rechtstheorie 1971, 171, 172.

fenden generalisierenden oder typisierenden Feststellungen".[65] Solche Feststellungen müssen nach anderen Regeln erfolgen als Einzelfallfeststellungen,[66] so dass die richterliche Tatsachenermittlung von der Evidenzbasierung oft nur mittelbar betroffen sein wird.[67] Deshalb liegt der Einwand nahe, das Recht sei für eine Evidenzbasierung zu einzelfallbezogen.[68] Doch kann man ernsthaft behaupten, das Recht sei einzelfallbezogener als die Medizin? Vielmehr eint doch „Juristen und Mediziner ein Denken, das notwendig den Einzelfall zum Ausgangspunkt hat",[69] aber beim Einzelfall auch nicht stehenbleiben kann. Denn *dass* Juristen häufig allgemeine Tatsachenbehauptungen aufstellen, wird schon durch die Existenz prozessualer „Erfahrungssätze" belegt und lässt sich auch sonst kaum bestreiten. „Werden sie aber eingesetzt, so müssen sie sich auch als solche kritisieren lassen".[70] Wenngleich also Evidenzbasierung nicht unmittelbar bei der Tatsachenermittlung im Einzelfall ansetzt, ist sie deshalb für die Jurisprudenz nicht unbedingt weniger relevant.

Desweiteren könnte einer Evidenzbasierung der Jurisprudenz entgegengehalten werden, dass ihr empirischer Grundansatz für Normwissenschaften schlicht abwegig sei.[71] Jedoch:

> „Praxiswissenschaften sind eo ipso auch Normwissenschaften. So gut wie jede medizinische Feststellung enthält neben einer Handlungsanmutung gleichzeitig auch eine Wertstellung".[72]

Evidenzbasierung scheint also keineswegs unvereinbar mit normativer Wissenschaft; darauf gehe ich gleich noch näher ein (unter B.).

---

[65] *Heldrich*, AcP 1986, 74, 80.

[66] Ausf. *Monahan/Walker*, Wis L Rev 1991, 569, 570 f. m.w.N. in Fn. 4, mit einer Unterscheidung zwischen "social authority" und "social fact".

[67] Das bedeutet bspw.: Ob ein Ereignis, auf das es im Gerichtsprozess ankommt, stattgefunden hat, bestimmt sich im Zivilrecht nach Beweislastregeln und im Strafrecht nach dem Zweifelssatz; ob einem Zeugen zu glauben ist, ist nur insoweit eine Frage der Evidenzbasierung, als die vom Richter herangezogene Vernehmungslehre evidenzbasiert sein sollte.

[68] *Rachlinski*, Cornell L Rev 2011, 901, 902; *Lawless/Robbennolt/Ulen*, Methods 2010, 14: "Legal analysis tends to be directed at […] the details of a particular case. In contrast, empirical analysis is concerned with examining patterns in the aggregate."

[69] *Augsberg*, GesR 2012, 595, 599; *Reilly*, Brit Med J 2004, 991, 992: "In a very real sense, all health care is local"; diff. *Windeler*, Gesundheitswesen 2008, 418, 418: „Das Grundkonzept einer evidenzbasierten Medizin sollte sinnvollerweise sowohl auf der Ebene der Versorgung individueller Patienten als auch bei Entscheidungen auf Systemebene Anwendung finden."

[70] *Jost*, Soziologische Feststellungen 1979, 161.

[71] So der Leitartikel *o.V.*, Can Med Assoc J 1999, 229 über "evidence-based morality".

[72] *Raspe*, GesR 2011, 449; *Augsberg*, GesR 2012, 595, 599 m.w.N.; *Reilly*, Brit Med J 2004, 991, 992: "in medicine good science is not the sole determinant of the right things to do. Political, economic, and sociocultural considerations sometimes trump the scientific ones."

Schließlich wird auch der diffuse Vorwurf erhoben, der „empiristische Ton“ in der Rechtswissenschaft entspreche einer „eigentlich vormodernen Ontologie“, die aus den Lehren des postmodernen Konstruktivismus keine „spürbare Konsequenz“ zöge.[73] Unklar bleibt, welche Konsequenz das sein könnte – jedenfalls aber nicht der *Verzicht* auf die „Betrachtung von außerrechtlichen Erkenntnisverfahren und einen reflektierteren Umgang mit fachfremdem Wissen“.[74] Immerhin gilt unter Fachfremden schon die bloße Frage, ob Recht evidenzbasiert sein müsse, als „ziemlich einfältig“,[75] da offensichtlich zu bejahen.

Versucht man also eine Zusammenführung der beiden Konzepte, so lässt sich „evidenzbasierte Jurisprudenz“ mit Blick auf die oben (I.) entwickelte Begriffsdeutung als Integration von zwei Säulen verstehen – von juristischer Lebenserfahrung und empirischer Erkenntnis. Nur die zweite Säule, das verdeutlicht der Untertitel dieser Arbeit, wird Gegenstand der folgenden Darstellung sein. Dass ich dadurch das Programm der evidenzbasierten Jurisprudenz nur halb einlöse, meine ich damit rechtfertigen zu können, dass die andere Hälfte – die „Sachkenntnis der Juristen […] in ihrer amorphen Erfahrungssättigung“[76] – ohnehin versiert gehandhabt wird. Kritische Worte über diese „Erfahrungssättigung“ werden zwar nicht ausbleiben, aber sie dienen ganz sicher nicht dazu, die juristische Lebenserfahrung als „Hosenbodenurteile“,[77] „Sesselempirismus“[78] oder bloße „Spekulation“[79] zu brandmarken und als „methodisch unzureichende *anecdotal evidence* [zu] belächel[n]“.[80] Vielmehr dient die Feststellung, dass Lebenserfahrung anekdotische Evidenz *ist* und ihre Beschränkungen *hat* (unten § 2 A.I.1.), nur der Einsicht, dass das Gebäude der evidenzbasierten Jurisprudenz ohne seine zweite Säule, die empirische Erkenntnis, unsicher steht. Dass auch diese Säule das Gebäude nicht allein tragen kann, versteht sich von selbst.

Nun habe ich ausgehend von der Abgrenzung zweier Verständnisse von „Evidenz“ und einer Erläuterung des Adjektivs „evidenzbasiert“ dessen Verwendung in der Medizin nachverfolgt, seine Verfestigung als Rechtsbegriff dargestellt und einige Parallelen zwischen Medizin und Recht aufgezeigt, die eine Urbarmachung dieses Topos’ für die Jurisprudenz empfehlen. Daraus habe ich das Konzept einer „evidenzbasierten Jurisprudenz“

[73] *Augsberg*, Staat 2012, 117 f.

[74] So auch *Augsberg*, Staat 2012, 117, 121.

[75] So der provokante letzte Satz im Leitartikel von *Carr-Hill*, Soc Sci Med 1995, 1467, 1468: “indeed, would it not been [sic] seen as rather silly to promote ‘evidence-based law’?”

[76] *Schön* in: Engel/Schön, Proprium 2007, 313, 315.

[77] *Zeisel/Kaye*, Figures 1997, 1: “relying solely on seat-of-the-pants judgments.”

[78] *Rhode*, Harv L Rev 2002, 1327, 1343 über “the limitations of armchair empiricism”.

[79] *Wozner*, Cornell L Rev 2011, 925, 927: “judges do not rely on empirical research but on speculation.”; zum Wert unverhohlener Spekulation unten § 1 E.III.

[80] *Schön* in: Engel/Schön, Proprium 2007, 313, 315.

entwickelt als Integration von Lebenserfahrung und empirischer Erkenntnis. Zweck dieser evidenzbasierten Jurisprudenz ist die seit jeher

> „genuine Aufgabe der Jurisprudenz, festzustellen, welche außerrechtlichen Fakten für eine bestimmte normative Fragestellung herangezogen werden sollen und ob die Ergebnisse der Nachbarwissenschaften dafür sich als fruchtbar erweisen.“[81]

Das bedeutet allerdings ebenso wenig eine Bejahung der eng verstandenen Frage, ob die Rechtswissenschaft eine empirische „Wende“ benötige,[82] wie die evidenzbasierte Medizin ihrerseits eine echte „empirische Wende“ herbeigeführt hat.[83] In beiden Disziplinen ist letztlich kein Paradigmenwechsel erforderlich, sondern eine Erweiterung ihrer Perspektiven – die aber gründlich, konsequent und reflektiert.

III. Gang der weiteren Darstellung

Die Erweiterung der juristischen Perspektive durch empirische Methoden – und deren Erkenntniswert – untersuche ich in der vorliegenden Arbeit in fünf Kapiteln, die aus einem Einführungskapitel, zwei Kapiteln zu den Grundlagen empirischer Forschung, einem Kapitel mit ausführlichem Anwendungsbeispiel und einem Schlusskapitel bestehen.

Im Rest dieses Einführungskapitels werde ich das Verhältnis von empirischer und juristischer Forschung untersuchen, deren jeweilige Rollen und Schnittstellen aufzeigen und wichtige Grundbegriffe klären. Abschließend werde ich die empirische Forschung konkret auf das Gesellschaftsrecht beziehen und die beiden Hauptrichtungen empirischer Gesellschaftsrechtsforschung in Deutschland und den USA charakterisieren.

Die beiden Grundlagenkapitel (§ 2 Grundlegung einer pragmatischen Rezeptionslehre und § 3 Die Gültigkeiten empirischer Forschung) stehen im Verhältnis der Spezialität zueinander: Das erste gibt die Grundstruktur vor, das zweite betrachtet einen wichtigen Teil dieser Struktur vertieft.

Im ersten der beiden Kapitel werde ich die quantitativ-empirische Methodik einerseits *beschreiben*, andererseits aber auch *systematisieren* und *kritisch reflektieren*. Dazu werde ich sieben „Rezeptionsregeln“ entwerfen, die dabei helfen sollen, empirische Erkenntnisse zu würdigen und zugleich ihre Beschränkungen zu verstehen.

Im zweiten Grundlagenkapitel werde ich herausarbeiten, wie innerhalb dieser Grundstrukturen nach der sog. Gültigkeitenlehre verschiedene Erhebungsmethoden unterschieden und kritisiert werden können. Um das Verständnis zu erleichtern, werde ich Beispielstudien mit gesellschaftsrechtli-

---

[81] *Schön* in: Engel/Schön, Proprium 2007, 313, 318; ebenso unten bei Fn. 198.
[82] Titelgebend bei *Petersen*, Staat 2010, 435.
[83] Missverständlich deshalb *Raspe*, GesR 2012, 584, 585.

chem Anwendungsbezug darstellen; als roter Faden soll mir die Frage dienen, ob höhere Erfolgsvergütung dazu geeignet ist, die Leistung von Vorstandsmitgliedern zu erhöhen.

Im Anwendungskapitel (§ 4 Hält das Kollegialprinzip, was es verspricht?) werde ich das sog. Kollegialprinzip untersuchen, das sich prominent im Gesellschaftsrecht, aber auch in anderen Rechtsgebieten findet. Ich werde herausarbeiten, dass und auf welchen empirischen Annahmen die Geltung des Kollegialprinzips beruht. Danach werde ich die empirische Forschung sichten und diese Annahmen einer Überprüfung unterziehen. Am Ende schließe ich den Kreis und erörtere mögliche rechtsdogmatische und rechtspolitische Folgerungen dieser empirischen Überprüfung.

Im Schlusskapitel (§ 5) fasse ich die Erkenntnisse der Arbeit in Thesen zusammen.

B. Recht und Empirie

Es ist das Verdienst des schottischen Philosophen *David Hume*, auf die erkenntnistheoretische Kluft zwischen Sein und Sollen hingewiesen zu haben.[84] Wissenschaftliche Fragestellungen werden seither oft in zwei Klassen eingeteilt:

Die einen sind ontologisch, positiv, deskriptiv – es geht um eine Lehre vom Sein (griech. ὄν *on* und λόγος *logos*), die Feststehendes (frz. *positif*) beschreibt (lat. *describere*).

Die anderen sind deontologisch, normativ, präskriptiv – es geht um eine Lehre vom Sollen (griech. δέον *deon* und λόγος *logos*), die Maßgebliches (frz. *normatif*) vorschreibt (lat. *praescribere*).

Je nachdem, wie man Wesen und Funktion des Rechts versteht, lässt sich die Wissenschaft vom Recht eher anhand der einen oder eher anhand der anderen Art von Fragestellungen betreiben. Dem entsprechen zwei Blickwinkel, die als Innenperspektive des Teilnehmers bzw. als Außenperspektive des Beobachters charakterisiert werden können.[85] Die stärkste Innenperspektive haben die radikal positivistischen Philosophien des Rechts. Sie halten das Recht für eine in sich geschlossene Sollensordnung und sehen die Aufgabe der Rechtswissenschaft darin, aus den Axiomen des Rechts mittels kanonischer Methoden der Begriffslogik abzuleiten, was

---

[84] *Hume*, Treatise 1739, T3.1.1.27; abl. *Upmeier*, Fakten 2010, 38 f. m.Verw. auf Vorarbeiten Spinozas.

[85] Ausf. *Mastronardi*, Juristisches Denken 2003, Rn. 639 ff. (Kap. III.2); *Albert* in: Vernunft 2011, 163, 171: „eine *realistisch-soziologische* und eine *normativistisch-analytische* Richtung".

sein *soll*, ohne Rücksicht darauf, was *ist*.[86] Die stärkste Außenperspektive hingegen hat der Rechtsrealismus, der „im Recht nicht eine ideelle Sollensordnung, sondern eine Wirklichkeit“ sieht: „Recht wird als Sozialtechnologie (*social engineering*) aufgefasst, als Instrument für die Machbarkeit gesellschaftlichen Wandels.“[87] Zwischen beiden Extremen liegt ein Spektrum an Philosophien, die Innen- und Außenperspektive miteinander verbinden und deshalb Sein und Sollen – Empirie und Dogmatik[88] – in unterschiedlicher Akzentuierung heranziehen.

Es gehört zu den ebenso bemerkenswerten wie folgenreichen Zufällen der Geschichte, dass sich die Rechtsphilosophien beiderseits des Atlantik auf unterschiedliche Extreme des Spektrums zu bewegt haben:[89] Während Rechtswissenschaft in den USA als angewandte Sozialwissenschaft im Sinne des Rechtsrealismus begriffen wird, dominierten in europäischen Rechtssystemen stets die stärker dogmatischen Schulen der Begriffs-, Interessen- bzw. Wertungsjurisprudenz. Daraus erklärt sich nicht nur die unterschiedliche Verbreitung und Akzeptanz der Rechtsökonomik,[90] sondern auch das unterschiedliche Bemühen um die empirische Fundierung des Rechts.[91] „Der amerikanische Pragmatismus war ein Wegbereiter“ dafür, dass „die empirische Rechtsforschung in Amerika vielleicht etwas rascher vorangeschritten ist als anderswo“[92] und sich zunehmend in unterschiedliche Schulen mit eigenen Methoden ausdifferenziert.[93]

---

[86] *Starck*, JZ 1972, 609, 610 f. zu den Theorien *Paul Labands* und *Hans Kelsens.*

[87] *Mastronardi*, Juristisches Denken 2003, Rn. 704 f.; *Starck*, JZ 1972, 609, 611 zum Freirechtsdenken *Theodor Geigers*; vgl. auch *Upmeier*, Fakten 2010, 36: „Ein besonders radikaler Versuch in dieser Richtung ist der ‚skandinavische Rechtsrealismus‘.“

[88] „Empirie“ stammt von griech. εμπειρία *empeiria*, Erfahrung und geht zurück auf das urindogermanische Wort für „versuchen“, „Dogmatik“ stammt von griech. δόγμα *dógma*, Meinung / Lehrsatz und geht zurück auf das urindogermanische Wort für „annehmen“ (*Harper*, Etymology 2014, Stichworte “empiric”, “dogma” und “decent”).

[89] Ausf. *Grechenig/Gelter*, RabelsZ 2008, 513 = *Grechenig/Gelter*, Hastings Int Comp L Rev 2008, 295.

[90] So das Ergebnis der instruktiven Untersuchung von *Grechenig* und *Gelter* (vorige Fn.).

[91] *Diamond*, U Ill L Rev 2002, 803, 822: “The legal realism movement provided the first significant and visible forum for the intersection between applied social science and legal scholarship […] its empirical vein developed largely by such professors as Underhill Moore, Charles Clark and William Douglas”; ebenso *Nard*, Wake Forest L Rev 1995, 347, 359 m.w.N. in Fn. 6; *Baldwin/Davis* in: Cane/Tushnet, Hdb Legal Studies 2003, 880, 882; *Kritzer* in: Cane/Kritzer, Hdb Empirical 2010, 875, 883: “there was a strong linkage between the legal realism movement in the United States and early empirical research on law”, wenn auch mit unklarer Ursachenrichtung (ebd. 879).

[92] *Zeisel*, JZ 1974, 561; *Riegert*, Dick Int L Ann 1983, 1, 4: “By way of comparison, there is probably more social-science research related to law in the United States than in Germany”.

[93] Einen aktuellen Überblick bieten *Suchman/Mertz*, Ann Rev L Soc Sci 2010, 555 sowie die Beiträge zum Symposium für Professor *Thomas Ulen* (18.–20.11.2010), etwa *Cooter*, U Ill L Rev 2011, 1475; *Eisenberg*, U Ill L Rev 2011, 1713; *Heise*, U Ill L Rev 2011, 1739.

Dass sich in Deutschland heute ein ganz anderes Bild bietet, ist nur praktisch eine zwingende Folge des deutschen Rechtsdogmatismus: *Zwingend* insoweit, als das Rechtsstudium angehende Juristen in Deutschland gar nicht mit den Kenntnissen ausstattet, die zu ernsthafter empirischer Forschung erforderlich sind.[94] *Nur praktisch* insoweit, als die in Deutschland vorherrschende Rechtstheorie gerade keine radikal positivistische ist.[95] Das mittlerweile sprichwörtliche „Hin- und Herwandern des Blickes zwischen Obersatz und Lebenssachverhalt“[96] soll ja gerade die starre begriffslogische Deduktion durch eine hermeneutische Spirale ersetzen, aus der auch der Obersatz nicht unverändert hervorgehen kann.[97] In einem gewissen Umfang *bedarf* die Rechtswissenschaft sogar der Empirie:

> „Beobachten und Teilnehmen gehören [.] beide zum juristischen Denken. Doch es gibt eine Gewichtung: Der Jurist ist primär Teilnehmer. Das Beobachten ist die notwendige Ergänzung zu seinem Tagesgeschäft.“[98]

> „Folglich muß die *Rechtsdogmatik*, die die Rechtsnormen zum Zwecke der Anwendung wissenschaftlich aufbereitet, Wirklichkeit berücksichtigen und die *juristische Methodenlehre* das Problem der Empirie in Betracht ziehen.“[99]

Für diese Einbeziehung von Empirie in die juristische Methodenlehre tritt in Deutschland vor allem die Strukturierende Rechtslehre ein, die eine „Normbereichsanalyse als wesentlichen Faktor juristischer Entscheidung“ begreift, also rechtsrelevante Tatsachen „*empirisch* zu ermitteln und mit den Elementen der Normtextauslegung rational, d. h. […] systematisch“ zu verbinden versucht.[100] Doch auch in der herrschenden Rechtsmethodik, die so weit nicht geht, ist „jedenfalls im Ergebnis nahezu unbestritten“, dass Wirklichkeitserkenntnis „in der Rechtswissenschaft insgesamt wie auch und gerade in deren dogmatischem Zweig ihren festen Platz einnimmt“.[101] Aus rechtstheoretischen Gründen spricht deshalb nichts gegen eine Evidenzbasierung im Recht, weil und solange sie normative Bewertungen nicht

---

[94] Dazu noch unten bei Fn. 179.

[95] Vgl. *Upmeier*, Fakten 2010, 15 f.: die tatsächliche „Zurückhaltung“ gegenüber „statistischen Untersuchungen“ sei Folge der Ausbildung, aber nicht mit dem „Rechtssystem *zwangsläufig* verbunden“.

[96] *Engisch*, Gesetzesanwendung 1963, 15; ähnl. *Mastronardi*, Juristisches Denken 2003, Rn. 245: „Juristisches Denken enthält immer beides: Sachverhalt und Norm, Beschreibung und Bewertung.“; deutlicher noch Rn. 285.

[97] Zur hermeneutischen Methode nur *Mastronardi*, Juristisches Denken 2003, Rn. 90 ff.

[98] *Mastronardi*, Juristisches Denken 2003, Rn. 54.

[99] *Starck*, JZ 1972, 609, 614 (These 2).

[100] *Müller/Christensen*, Methodik I 2013, 527 f. (Rn. 482 f.).

[101] *Gusy*, JZ 1991, 213; ähnl. prägnant *Schneider/Teitelbaum*, Utah L Rev 2006, 53, 57: “We will not trek through the long argument for empirical scholarship in legal studies. That has been done, admirably and often, and the case’s merits are now plain and familiar. We cannot debate them because no one will mount a systematic argument against them.”

ersetzen, sondern empirisch unterfüttern soll[102] – eben „auf Evidenz basieren". Umgekehrt lässt sich formulieren, dass die empirische Rechtsforschung dem naturalistischen Fehlschluss ohne Weiteres entgehen kann, indem sie „*konditionale Empfehlungen* […] auf der Basis einer als angestrebt unterstellten Zielsetzung" abgibt.[103]

Tiefer möchte ich auf die philosophische Verschränkung von Recht und Empirie und insbesondere auf die Ansicht, dass „Rechtswissenschaft als Realwissenschaft […] dringend notwendig" sei,[104] nicht eingehen. Der Rest dieser Schrift geht vielmehr davon aus, dass ein Mehr an wissenschaftlich fundiertem Verständnis der Rechtswirklichkeit besser ist als weniger oder gar keines,[105] und setzt dort an, wo Juristen bereits ausdrücklich ein Bedürfnis nach empirischen Erkenntnissen artikulieren. Die Untersuchung behandelt also aus dem Blickwinkel der empirischen Methodik „die bislang weitgehend offen gebliebene Frage nach dem ‚Wie' der Einbeziehung von Wirklichkeit", die andernorts bereits „vom Recht und der dogmatischen Rechtswissenschaft her angegangen" wurde.[106]

Eine pragmatische Darstellung dieser Methodik und ein Transfer bereits gewonnener Erkenntnisse können wohl am ehesten das praktische Potential der empirischen Rechtsforschung bewerten helfen.

---

[102] *Bilz*, JITE 2010, 194, 196: "empirical findings get used as one argument among many, on both sides of any given policy dispute."; ähnl. *Schneider/Teitelbaum*, Utah L Rev 2006, 53, 68 f. m.w.N.; schon *Jost*, Soziologische Feststellungen 1979, 22 nahm an, „daß Aussagen über Fakten für das Werturteil bedeutsam sind, nur erschöpft es sich nicht in ihnen."

[103] *Eidenmüller*, JZ 1999, 53, 54, 55; *Rachlinski*, Cornell L Rev 2011, 901, 918 f.; ähnl. *Alexy*, Argumentation 1983, 287; *Winter*, Rechtstheorie 1971, 171, 173 fragt, „ob Tatsachenurteile die Rationalität einer Auswahl zwischen verschiedenen Normvorschlägen erhöhen".

[104] So *Eidenmüller*, JZ 1999, 53, 61.

[105] *Eisenberg*, U Ill L Rev 2011, 1713, 1720: "a core principle seems indisputable: it is better to have more systematic knowledge of how the legal system works rather than less, regardless of the normative implications of that knowledge."; *Goldsmith/Vermeule*, U Chi L Rev 2002, 153, 165: "On the case level, courts might prefer to navigate with an inaccurate map than with no map at all."; *Hogarth*, J Acc Res 1982, S108, 113: "A little knowledge is better than none. […] We must educate ourselves not to expect too much of experiments but bear in mind that the possession of knowledge is relative. As stated by Erasmus, 'In the land of the blind, the one-eyed man is king.'"

[106] *Gusy*, JZ 1991, 213, 214.

C. Empirie als Methode

Im akademischen Alltag drängen deutsche Rechtswissenschaftler selten auf die empirische Unterfütterung ihrer Arbeit. Gleichwohl bricht ein solches Bedürfnis bisweilen durch. So herrschte 2006 beim Hamburger Symposium zu den Zukunftsperspektiven der privatrechtlichen Forschung

> „über die Grundtendenzen weitgehende Übereinstimmung. Sämtliche Referenten betonten die Notwendigkeit interdisziplinärer Forschung […] Empirischen Ansätzen solle ein wachsender Stellenwert zukommen.“[107]

Da es aus wissenschaftstheoretischer Sicht keine Disziplin gibt, die rein empirisch tätig ist,[108] umfasst die zitierte Forderung genaugenommen zwei verschiedene Aspekte: *Materiell* die Zusammenarbeit mit anderen Disziplinen, *methodisch* die empirische Herangehensweise.[109] Wenn aber im deutschen Rechtsraum von „empirischer Rechtsforschung“ die Rede ist, war bisher stets die materielle Zusammenarbeit mit der Soziologie gemeint,[110] die von jeher auch (und vielleicht sogar vorrangig) auf einer theoretischen Ebene stattfand.[111] Dass Juristen grenzüberschreitende Untersuchungsgegenstände bevorzugt aus der materiellen Sicht *einer* Partnerdisziplin umreißen, verdeutlichen neben der Rechtssoziologie (*socio-legal research* bzw. *law and society*)[112] auch die Rechtspsychologie (*law and psychology*),[113] die Rechtsethologie (*law and biology*),[114] die Rechtsökonomik (*law*

---

[107] *Hopt/Trautmann*, ZGR 2007, 480, 483.

[108] Vgl. schon *Starck*, JZ 1972, 609, 613; ausf. *Ulen*, U Ill L Rev 2002, 875, 882 ff.

[109] *Neumann/Krieger*, Clin L Rev 2003, 349, 353: “To law faculty, empirical work is interdisciplinary […]. But much interdisciplinary work is not empirical.”; *Eidenmüller*, ZGR 2007, 484, 494 = *Eidenmüller*, JZ 2007, 487, 491 bezeichnet die „interdisziplinäre Öffnung“ sogar als „eine andere methodische Entwicklung“ als „die Hinwendung zur Empirie“.

[110] *Röhl*, RSozBlog 2013: „Die Rechtssoziologie weiß sich daher nur noch mit einem imperialistischen Begriff ihres Faches zu helfen, der alle empirisch orientierte Rechtsforschung einschließt.“

[111] So etwa *Strempel* in: Broda u.a., FS Wassermann 1985, 223: „Da empirische Rechtsforschung den mehr praktischen Aspekt der Rechtstatsachenforschung und den mehr theoretischen Aspekt der Rechtssoziologie umfasst, […]“; ähnl. *Heldrich*, AcP 1986, 74, 78.

[112] Für das Gesellschaftsrecht vgl. schon *Geiler* in: Predari u.a., Beiträge 1927, 593, 599 ff.; später v.a. befördert durch *Thomas Raiser*, Organisation 1969, auch im Sammelband *Raiser*, Rechtssoziologie 2011 zur Soziologie der Handelsgesellschaften (175) und zu Konflikten in Wirtschaftsunternehmen (207).

[113] Ausgewählte Beiträge aus den Jahren 1907–1986 in *Jakob/Rehbinder*, Rechtspsychologie 1987; aus zivilrechtlicher Sicht etwa *Rachlinski*, Cornell L Rev 1999, 739; *Rachlinski*, U Ill L Rev 2011, 1675.

[114] Dazu allg. *Gruter*, Verhaltensforschung 1976; *Schmidt*, Verhaltensforschung und Recht 1982; *Gruter*, Rechtsverhalten 1993; *Hof*, Rechtsethologie 1996; *Hof* in: Brand/Strempel, FS Blankenburg 1998, 65; *Elliott*, St Louis U L J 1997, 595; *Von Rohr*, Grundlagen 2001 sowie die Beiträge im Sammelband *Haft/Hof/Wesche*, Verhaltenstheorie 2001 und die Nachw. in

and economics)[115] und die Verhaltensökonomik des Rechts (*behavioral law and economics*).[116] Oft bezwecken und erreichen solche Querschnittsdisziplinen eine theoretische Integration juristischer Fragen auch ohne die empirische Herangehensweise,[117] wenngleich diese teilweise sehr prominent ausgeprägt sein mag.[118] Dass die vorliegende Arbeit ihren Gegenstand dagegen *ausschließlich* an der empirischen Methode statt an der Materie einer Partnerdisziplin ausrichtet, ist ungewöhnlich und bedarf daher besonderer Begründung.

Die Unterscheidung zwischen Materie und Methode findet sich schon innerhalb des Rechts, beispielsweise zwischen dem *Wirtschaftsrecht* als einer Rechtsmaterie und der *Auslegung* als einer rechtgebietsübergreifenden Methode. Ganz ähnlich – wenn auch auf anderer Ebene – lässt sich die *inter*disziplinär verstandene *Materie* der „empirischen Rechtsforschung“ unterscheiden von den *trans*disziplinären (fachgebietsübergreifenden) *Methoden* der „empirischen Sozialforschung“. Diese Methoden finden sich gleichermaßen in so unterschiedlichen Disziplinen wie Soziologie, Politikwissenschaft, Ökonomik, Betriebswirtschaftslehre, Psychologie, Pädagogik, Humanmedizin, Geographie, Geschichtswissenschaft, Ethnologie und – Rechtswissenschaft.[119] Obwohl zwischen diesen Disziplinen „enorme Unterschiede schon in den Denkrichtungen […] bestehen“, unterwerfen sich alle „demselben Prozedere bei einer Anwendung der empirischen Forschung“.[120] Dieses fachgebietsunabhängige „Prozedere“ ist es auch, das der

---

Heldrich, AcP 1986, 74, 79 Fn. 21.

[115] Statt aller *Eidenmüller*, Effizienz 2005; *Schäfer/Ott*, Lehrbuch 2012.

[116] Bahnbrechend *Jolls/Sunstein/Thaler*, Stan L Rev 1998, 1471; ausf. *Engel/Englerth u.a.*, Recht und Verhalten 2007.

[117] Für die USA moniert *Rachlinski*, Cornell L Rev 2011, 901, 906, dass viele Universitäten “had economically trained or economically oriented faculty for many years, of course, but few of these have also been empiricists.”

[118] Z.B. *Rachlinski*, Cornell L Rev 2011, 901, 910: “Behavioral law and economics is, at its core, empirical; its very reason for existence is the belief that careful empirical work will show that human beings often do not conform to the assumptions of microeconomic theory.”; ähnl. *Englerth* in: Engel u.a., Recht und Verhalten 2007, 60, 103 über „Behavioral Law and Economics [… und] seine spezifische empirische Forschungsmethode“.

[119] So *Diekmann*, Sozialforschung 2012, 20 ff., mit Bsp. aus jeder dieser Disziplinen; ebenso *Eisenberg* (unten Fn. 145) und *Epstein/Martin* in: Cane/Kritzer, Hdb Empirical 2010, 901, 902: “empirical research in law has methodological concerns that overlap with those in Biology, Chemistry, Economics, Medicine and Public Health, Political Science, Psychology, and Sociology”; ähnl. *Arlen/Talley*, Experimental 2008, li: experimentelle Erkenntnisse “transcend the legal setting, and […] are relevant to economists, psychologists and policy-makers interested in human behavior.”

[120] *Arndt*, Empirie 2008, 11; *Lawless/Robbennolt/Ulen*, Methods 2010, 2: “there is a common set of skills that runs through the empirical techniques that each academic discipline uses.”; a.A. für die Rechtstatsachenforschung *Gilles* in: Prütting/Rüssmann, FS Lüke 1997,

evidenzbasierten Medizin zugrunde liegt, ohne dass sie überhaupt materielle Theorien entwickelt.[121] Auch eine neuere Schule der empirischen Rechtsforschung in den USA (*empirical legal studies*) setzt die empirische Methode ohne Rückgriff auf ein bestimmtes disziplinäres Vorverständnis ein.[122] Man mag darin eine neue Religion sehen[123] oder ein schlicht pragmatisches Vorgehen, denn für die Zwecke der Evidenzbasierung kommt es schlicht „nicht auf den Nachweis der Bedeutung einer bestimmten Sozialwissenschaft" an.[124]

Es liegt mir fern zu leugnen, dass Grundlagenforschung innerhalb der (materiellen) Paradigmen *einer* Nachbarwissenschaft enormen Erkenntnisfortschritt über das Recht ermöglicht; man denke nur an die ökonomische Analyse gerade des Gesellschaftsrechts.[125] Gleichwohl birgt jede Forschung, die sich an Disziplingrenzen orientiert, Tücken.[126] Die Kanonisierung der akademischen Disziplinen suggeriert eine trennscharfe Abgrenzbarkeit ihrer Gegenstandsbereiche, die es nach der Natur der Sache nicht geben kann.[127] Disziplingrenzen sind weniger den Eigenheiten der zu untersuchenden Lebenswirklichkeit geschuldet als vielmehr historischen Pfadabhängigkeiten und der praktischen Handhabbarkeit des akademischen Lehr- und Forschungsbetriebs. Mit fortschreitendem Erkenntnisstand verändern sich diese Grenzen, entstehen neue Disziplinen und vergehen alte.[128] Daher sind „interdisziplinär" arbeitende Juristen zum einen der

---

139, 147.

[121] *Raspe*, GesR 2012, 584, 588 m.w.N. zur dagegen erhobenen Kritik.

[122] *Baldwin/Davis* in: Cane/Tushnet, Hdb Legal Studies 2003, 880, 893: "Few empirical researchers in law make any claim that their research is located within some overarching theoretical framework derived from one of the core social science disciplines."; ausf. *Eisenberg*, San Diego L Rev 2004, 1741; *Eisenberg*, U Ill L Rev 2011, 1713; *Chambliss*, L Ctmp Prob 2008, 17, 22 ff.; *Suchman/Mertz*, Ann Rev L Soc Sci 2010, 555; vgl. auch *Petersen*, Staat 2010, 435, 436 m.w.N. in Fn. 3.

[123] *Wozner*, Cornell L Rev 2011, 925, 926: "dramatic rise of the new religion of empiricism in legal studies in American law schools […] a trend very similar to that in Israel".

[124] So schon *Jost*, Soziologische Feststellungen 1979, 15.

[125] Früh schon *Steinitzer*, Theorie der AG 1908; später *Coase*, Economica 1937, 386; aus neuerer Zeit *Fleischer/Zimmer*, Effizienz 2008; zu einzelnen Theorien noch unten Fn. 133.

[126] Schilderung bei *Gessner*, JZ 1971, 324, 325 ff.

[127] Vgl. schon *Jost*, Soziologische Feststellungen 1979, 15 f.; *Beutel*, Experimentelle Rechtswissenschaft 1971, 203: „die Anforderungen einer wissenschaftlichen Suche nach juridischen Gesetzen werden sich quer durch die gegenwärtige Klassifizierung der Wissenschaften selbst ziehen".

[128] Auch *Baer* in: Grundmann u.a., FS Juristische Fakultät 2010, 917 Fn. 2 merkt an, dass die „intradisziplinären Sortierungen in die ‚Fächer' [.] in ihrer Sinnhaftigkeit aus sehr unterschiedlichen Perspektiven ebenso in Frage gestellt werden können wie die Versuche, neue ‚Rechtsgebiete' zu konturieren. Rechtsgebiete spiegeln in ihrem Zuschnitt Interessen- und auch Lebenslagen."

Willkür ihrer Partnerdisziplin ausgesetzt, was die Definition von Gegenstandsbereich und Forschungstrends angeht. Vertreter der Volkswirtschaftslehre (Ökonomik) etwa sind dafür berüchtigt, in jüngerer Zeit alles zum Gegenstand ihrer Forschung gemacht zu haben, was ihren Methoden nur irgend zugänglich war (sog. *economic imperialism*).[129] Das verändert natürlich auch das Wesen der Rechtsökonomik: Ihre klassischen Themen werden vernachlässigt und geraten unter Rechtfertigungsdruck, ihr Gegenstand verliert an Schärfe. Zum anderen sind „interdisziplinär" arbeitende Juristen auf die Methoden ihrer Partnerdisziplin beschränkt. Erneut am Beispiel der Ökonomik: Diese hat im Bereich der experimentellen Forschung methodische Beschränkungen entwickelt, die unumstößlichen Glaubenssätzen gleichkommen,[130] und im Kontext der eigenen Disziplin ihre gute Berechtigung haben mögen; zugleich verstellen sie aber den Blick auf andere Themen und Probleme, die für Juristen besonders relevant sein könnten (vgl. § 3 nach Fn. 150).

Diese Abhängigkeit von einer Partnerdisziplin lässt sich verringern, indem nicht eine bestimmte Disziplin, sondern die disziplinübergreifende empirische *Methode* herangezogen wird, die als solche kein bestimmtes Vorverständnis erfordert.[131] Die vorverständnisfreie Handhabung der Empirie entspricht einem zutiefst pragmatischen Anliegen und dürfte jene Art von Revierkämpfen vermeiden, die früheren empirischen Grenzüberschreitungen der Jurisprudenz einen so schweren Stand verschafften.[132]

Andererseits beschränkt sie den Blick natürlich wiederum in dramatischer Weise: Eine Arbeit zur *empirischen* Gesellschaftsrechtsforschung

---

[129] *Engel/Schön* in: Engel/Schön, Proprium 2007, IX, XI; *Stigler*, Scand J Econ 1984, 301, 311: "economics is an imperial science: it has been aggressive in addressing central problems in a considerable number of neighboring social disciplines, and without any invitations."; *Lazear*, Q J Econ 2000, 99, 103: "The most aggressive economic imperialists aim to explain all social behavior by using the tools of economics."; diff. *Mäki*, Phil Soc Sci 2009, 351.

[130] *Loewenstein*, Econ J 1999, 25, 33: "In more than thirty years of productive research, E[xperimental]E[conomist]s have developed some extremely compelling experimental conventions. For example, EE methods of ensuring incentive compatibility, the discouragement of deception, and the practice of reporting methods in sufficient detail to allow for replication"; *Hertwig/Ortmann*, Behav Brain Sci 2001, 383, 384: "The experimental standards in psychology, by contrast, are comparatively laissez-faire, allowing for a wider range of practices."

[131] *Eisenberg*, U Ill L Rev 2011, 1713, 1732: "A strength of E[mpirical]L[egal]S[tudies] […], though also sometimes a potential weakness, is the absence of the need for loyalty to a particular social science discipline." und weiter 1737: "No need exists to decide whether economics, sociology, political science, or psychology has the most persuasive or accurate view of the world."

[132] Vgl. *Lüdemann* in: Engel u.a., Recht und Verhalten 2007, 1; prägnant *Tontrup* in: Engel/Schön, Proprium 2007, 192, 193: „Man gelangt eben nicht in die Höhle des Löwen, es sei denn als Schaf oder Grundlagenfach."

kann wenig beitragen zu den für das Verständnis „der Gesellschaft“ so wesentlichen Versuchen ihrer *theoretischen* Modellierung.[133] Diese Selbstbeschränkung nehme ich aber mit Blick darauf hin, dass empirische Ansätze für Juristen womöglich leichter zugänglich und bislang weniger gründlich aufbereitet sind als theoretische.

D. Bruchlinien in der Empirie

Innerhalb „der“ empirischen Forschung sind vor allem zwei Abgrenzungen etabliert, zu denen im Rahmen der Standortbestimmung Position bezogen werden soll, weil sie zu einer sachgerechten Beschränkung der vorliegenden Untersuchung beitragen könnten.

I. Qualitative und quantitative Empirie

Empirische Forscher unterscheiden grundsätzlich zwischen zwei verschiedenen Forschungsansätzen. Diese werden mit den Adjektiven „qualitativ“ und „quantitativ“ bezeichnet und gewöhnlich nach ihrem Kommunikationsmedium unterschieden: Der eine Ansatz arbeite mit Begriffen bzw. nichtnumerischen „Verbalisierungen“ (also Qualitäten, von lat. *qualis*, wie beschaffen), der andere dagegen mit numerischen „Messwerten“ (Quantitäten, von lat. *quantus*, wie viel).[134] Diese Begriffswahl führt allerdings leicht in die Irre, denn tatsächlich arbeitet die als qualitativ bezeichnete Forschung ebenso mit Zahlen wie die als quantitativ bezeichnete Forschung mit Begriffen arbeitet.[135] Ein besseres Unterscheidungskriterium ist

---

[133] Meilensteine etwa *Jensen/Meckling*, J Fin Econ 1976, 305 (*nexus of contracts*); *Tirole*, J L Econ Org 1986, 181 (*principal-agent-supervisor*); *Blair/Stout*, Va L Rev 1999, 247 (*team production*); grundsätzlich *Eidenmüller*, ZGR 2007, 484, 493 = *Eidenmüller*, JZ 2007, 487, 491: „Die Öffnung der Rechtswissenschaft für die verhaltensbezogenen Modelle und Erkenntnisse der ökonomischen Theorie und der Psychologie besitzt gerade im Unternehmensrecht eine hohe Bedeutung. Man denke etwa an den Stellenwert, den die principal-agent-Theorie für die Analyse von Interessenkonflikten innerhalb einer Kapitalgesellschaft gewonnen hat.“

[134] *Bortz/Döring*, Forschungsmethoden 2006, 296; *Epstein/King*, U Chi L Rev 2002, 1, 2: “numerical (quantitative) or nonnumerical (qualitative)”; ebenso *Chui* in: McConville/Chui, Methods 2007, 46, 48; zur Etymologie vgl. *Harper*, Etymology 2014, Stichworte “quality” / “quantity”; *Köbler*, Lateinisches Wb. 2009; Dudenredaktion, Fremdwörterbuch 2000, 1119 („Qualität“ und „Quantität“).

[135] Das wird etwa daran deutlich, dass der „qualitative“ Ansatz „kaum umhinkommt, zu quantifizieren“ (*Diekmann*, Sozialforschung 2012, 607 m.w.N.) und dass Daten der „quantitativen“ Forschung auf zwei ihrer vier Skalenniveaus (unten § 2 A.II.3.) nur verbal und nicht durch Zahlen darstellbar, also doch „qualitative Merkmale“ sind (*Petersen/Goerg* in: Towfigh/Petersen, Methoden 2010, 201, 213).

wohl die *Gleichförmigkeit* der erhobenen Daten: Während die als qualitativ bezeichnete Forschung Einzelfälle mit Blick auf ihre Strukturen und Eigenheiten untersucht, geht es der als quantitativ bezeichneten Forschung um Fallgruppen oder Fallgesamtheiten mit Blick auf ihre Gemeinsamkeiten und Beziehungen.[136] Daher wären die Begriffe „Einzelfall-“ und „Sammelforschung“ womöglich treffender als „qualitative“ und „quantitative“, allerdings behalte ich die etablierten Begriffe bei, um unnötige Verwirrung zu vermeiden.

Die Zweiteilung in quantitative und qualitative Forschung ist historisch gewachsen. Als in der Renaissance und Aufklärung das Interesse an empirischen Methoden aufkeimte,

> „entwickelten sich in Deutschland die *Universitätsstatistik* und in England die *politische Arithmetik*, zwei gegensätzliche Schulen, deren Traditionen heute als die beiden Wurzeln der empirischen Sozialforschung und Statistik gelten können.“[137]

Deshalb wird „empirische Sozialforschung“ bisweilen mit der qualitativen Forschung gleichgesetzt,[138] „Statistik“ hingegen mit der quantitativen. Erstere ist eine klassische Domäne der Soziologie und derjenigen empirischen Disziplinen, die sich in einer philosophisch-geisteswissenschaftlichen Tradition sehen, während letztere vor allem von Disziplinen betrieben wird, die sich an der naturwissenschaftlichen Tradition orientieren, wie etwa weite Teile der Psychologie und der empirischen Ökonomik.

So tief wie die Zweiteilung historisch verwurzelt ist, so wenig ist sie erkenntnistheoretisch zwingend. Obwohl sich qualitative und quantitative Forschung sowohl im Erkenntnisinteresse[139] als auch den Methoden der Datenerhebung[140] und Datenauswertung[141] unterscheiden, ist keiner der beiden Ansätze überlegen oder vorrangig:

---

[136] Ähnl. *Hussy/Schreier/Echterhoff*, Forschungsmethoden 2010, 9: „quantitativ“ sei standardisiertes Messen, „qualitativ“ unstandardisiertes Sinnverstehen.

[137] *Diekmann*, Sozialforschung 2012, 92 und weiter 95: „Die beiden gegensätzlichen Schulen befehdeten sich nicht schlecht“.

[138] Allerdings gerade nicht von *Diekmann* (vorige Fn.), der in seinem Lehrbuch „Empirische Sozialforschung“ auch die Statistik behandelt.

[139] Etwa *Baldwin/Davis* in: Cane/Tushnet, Hdb Legal Studies 2003, 880, 891 f.: “[in] qualitative research […] the aim is to improve understanding, not to gather evidence”.

[140] Vgl. aber *Röhl*, Rechtssoziologie 1987, 118: „Qualitative Sozialforschung bedient sich im Prinzip derselben Erhebungsmethoden wie wie quantitative. Sie verzichtet nur auf standardisierte Instrumente.“

[141] Z.B. *Goldsmith/Vermeule*, U Chi L Rev 2002, 153, 159: “the statistical worldview does not translate cleanly to the […] case study”; anders *Gorard*, Brit J Sociol Edu 2006, 67, 68: “the logic of statistical analysis is little different, except in terminology, from the analysis of other forms of data—especially those traditionally referred to as ‘qualitative’.”; ausf. *Neumann/Krieger*, Clin L Rev 2003, 349, 354 f. zum Streit um die von *Epstein/King*, U Chi L Rev 2002, 1 verfochtene Methodeneinheit.

„In der empirischen Sozialforschung geht es immer um inhaltliche Fragestellungen, um ‚Probleme' im ganz allgemeinen Sinne, die in der erwähnten Terminologie als ‚qualitative' Sachverhalte gelten müssen. […] Dabei können qualifizierende und quantifizierende Aspekte in verschiedenen Phasen des Forschungsprozesses mit unterschiedlichem Stellenwert einfließen, fast immer wird es aber eine Kombination beider Vorgehensweisen sein."[142]

Folglich ergänzen sich qualitative und quantitative Forschung gegenseitig;[143] beide sind notwendige Bestandteile jeder Evidenzbasierung.[144]

Da es allerdings den Rahmen der vorliegenden Arbeit sprengen würde, beide Bereiche halbwegs umfassend zu beleuchten und in ihrer Wechselbeziehung darzustellen, muss ich mich im Folgenden auf einen der beiden Ansätze beschränken. Die Wahl fällt dabei auf den quantitativen Ansatz, der sowohl die heutige empirische Rechtsforschung allgemein[145] als auch die empirische Forschung zum deutschen und US-amerikanischen Gesellschaftsrecht dominiert (dazu gleich G.). Das mag einer eigenwilligen Geisteshaltung der Rechtspraxis geschuldet sein,[146] soll hier aber nicht hinterfragt werden. Stattdessen beschränkt sich die weitere Darstellung in ihrer Breite auf quantitativ-empirische Forschung und blendet qualitative Methoden weitgehend aus.[147]

Diese Schwerpunktsetzung soll allerdings keinesfalls verleugnen, dass der qualitative Ansatz nach wie vor zeitgemäß ist und enormes Erkenntnis-

---

[142] *Diekmann*, Sozialforschung 2012, 607 m.w.N.; *Baldwin/Davis* in: Cane/Tushnet, Hdb Legal Studies 2003, 880, 892: "Increasingly, empirical legal researchers are employing a combination of qualitative and quantitative techniques, seeking to harness the strengths of both."

[143] Ausf. *Nielsen* in: Cane/Kritzer, Hdb Empirical 2010, 951 ff. (Kap. 39); *Hussy/Schreier/Echterhoff*, Forschungsmethoden 2010, 273 ff. (Teil III) zu sog. *mixed-methods*-Untersuchungen; ebenso *Lawless/Robbennolt/Ulen*, Methods 2010, 46 f.; *Arndt*, Empirie 2008, 17 (in Fn. 45 m.Verw. auf *Diekmann*, Sozialforschung 2012, 19) zur „Triangulation, also Kombination und Wiederholung der verschiedenen Methoden".

[144] So für die Medizin *Green/Britten*, Brit Med J 1998, 1230.

[145] *Kritzer* in: Cane/Kritzer, Hdb Empirical 2010, 875, 883: "contemporary empirical legal studies [… are] largely quantitative in character."; *Eisenberg*, U Ill L Rev 2011, 1713, 1719: "ELS employs […] usually, but not always, the methodology of statistical analysis [… taught by] economics, psychology, health care, policy, political science, criminology, finance, or sociology."; vgl. schon *Schuck*, J L Edu 1989, 323: "When I speak of 'empirical research,' I refer primarily to statistical studies […] in an effort to detect important regularities (or irregularities) that have not previously been identified or verified."

[146] So *Gessner* in: Plett/Ziegert, Empirische Rechtsforschung 1984, 69, 81 = *Gessner*, SozW 1984, 480, 484: „Juristen haben angesichts ihrer täglichen Aufgabe, in sozialen Problemfällen zu intervenieren, ein überaus festgefügtes Bild sozialer Zusammenhänge, das sie sich allenfalls durch die harten Fakten quantitativer Untersuchungen […] korrigieren lassen. Das ist angesichts spezifischer Aussagegrenzen aggregierter Daten sicherlich zu bedauern."

[147] Eine Beispielstudie wird allerdings unten in § 3 bei Fn. 258 erwähnt.

potential hat. Interessierte Leser seien deshalb auf einschlägige Lehrbuchdarstellungen verwiesen.[148]

II. Grundlagen- und Anwendungsforschung

Nur kurz anzureißen ist die weitere Abgrenzung zwischen grundlagen- und anwendungsbezogener Forschung, die zwar weit verbreitet ist, aber zugleich recht unscharf[149] und aus *methodischer* Sicht unergiebig, denn beide Ansätze unterscheiden sich in ihren Methoden überhaupt nicht, weswegen im Rahmen der vorliegenden Arbeit auch keine weitere Abschichtung nötig ist. Gleichwohl ist die Differenzierung in zweierlei Hinsicht konzeptionell aufschlussreich:

Zum einen hilft sie dabei, historische Entwicklungen und divergierende Erkenntnisinteressen zu verstehen. Das wird zum Beispiel am unterschiedlichen Verständnis von empirischer Rechtsforschung in Deutschland und den USA deutlich: Während in Deutschland darunter meist eine unmittelbar rechtspolitische Anwendungsforschung ohne theoretische Reflexion verstanden wird[150] – nicht selten sogar Auftragsforschung für Unternehmen oder Ministerien[151] – gilt sie in den USA als sozialwissenschaftliche Grundlagenforschung über die Bestimmungsgründe menschlichen Verhaltens.[152] Dementsprechend beherrschen zwei unterschiedliche Forschungsrichtungen mit speziellen Fragestellungen und Methoden die empirische Gesellschaftsrechtsforschung: In Deutschland die Rechtstatsachenforschung, in den USA die Ereignisstudienforschung (dazu gleich G.). Als

---

[148] Für Juristen sehr knapp *Lercher* in: Barta u.a., Rechtstatsachenforschung 2009, 205, 218 ff.; Überblicke bei *Bortz/Döring*, Forschungsmethoden 2006, 295 ff. (Kap. 5); *Sedlmeier/Renkewitz*, Forschungsmethoden 2008, 741 ff. (Kap. 25); *Hussy/Schreier/Echterhoff*, Forschungsmethoden 2010, 19 ff., 179 ff. (Kap. 5–8); ausf. dargestellt von *Lamnek*, Lehrbuch 2010; *Flick*, Einführung 2010; *Mayring*, Anleitung 2002.

[149] Auch *Diekmann*, Sozialforschung 2012, 37 stellt fest, dass „die Abgrenzung zwischen ‚reiner' Grundlagenforschung und angewandter Forschung nicht immer [.] eindeutig ist und wechselseitige Zusammenhänge bestehen".

[150] Angesichts der fehlenden „wissenschaftstheoretisch hinreichend abgesicherten Basis" der Rechtstatsachenforschung fordert etwa *Gilles* in: Prütting/Rüssmann, FS Lüke 1997, 139, 147 f., „dass man sich [.] sehr viel extensiver und intensiver als bisher mit ihren *theoretisch-methodologischen Grundfragen* befassen sollte."

[151] Vgl. *Strempel* in: Broda u.a., FS Wassermann 1985, 223; *Strempel* in: Schäffer, Rechtskultur 1987, 87.

[152] *Riegert*, Dick Int L Ann 1983, 1, 4: "a larger part of the American research is aimed at the accumulation of general social-science knowledge, rather than specifically at immediate law reform problems."; *Eidenmüller*, JZ 1999, 53, 54: „das Erkenntnisinteresse einer so [d.h. empirisch] betriebenen Rechtswissenschaft […] liegt in der Gewinnung *instrumentell verwertbaren Wissens* über kausale Zusammenhänge in der Wirklichkeit, deren Produkt Recht ist oder die von Recht ausgelöst werden."; vgl. auch *Arlen/Talley* (Fn. 119).

weiteres Beispiel für die Konsequenzen der Unterscheidung zwischen Grundlagen- und Anwendungsforschung werde ich in § 4 C.II.3. darstellen, wie sich ein akademischer Streit über vermeintlich grundstürzende neue Forschungsergebnisse bei näherem Hinsehen als bloßes Missverständnis zwischen Forschern aus dem Grundlagen- und dem Anwendungslager entpuppen kann.

Zum anderen hilft die Unterscheidung zwischen Grundlagen- und Anwendungsforschung bei der pragmatischen Rezeption empirischer Erkenntnisse, deren Grundlagen im nächsten Kapitel (§ 2) herauszuarbeiten sind. Denn beide Ansätze unterscheiden sich im Erkenntnisinteresse,[153] in ihren Gültigkeitsvoraussetzungen[154] und in ihren Entstehungsbedingungen – insbesondere dadurch,

> „dass Interessen, Ideologien und Wertvorstellungen von Auftraggebern und Sozialforschern einen erheblichen Einfluss auf die Forschungsresultate ausüben können […] wenn die Forschung anwendungsbezogen ist und sich das Forschungsinteresse auf kontroverse Themen richtet".[155]

Dabei spielen auch versteckte Motive eine Rolle: "Nicht selten [.] werden evaluierende Gutachten in Auftrag gegeben, um eine kontroverse Entscheidung hinauszuzögern oder eine längst getroffene Entscheidung mit einem ‚Gefälligkeitsgutachten' zu legitimieren."[156] Aus diesen Gründen sollte Auftragsforschung ebenso klar gekennzeichnet und der Auftraggeber benannt werden[157] wie Parteigutachten im juristischen Fachschrifttum.[158]

Die evidenzbasierte Medizin, an die sich die evidenzbasierte Jurisprudenz sprachlich anlehnt, wird weithin als Anwendungsforschung verstan-

---

[153] *Diekmann*, Sozialforschung 2012, 37: „Denkt man bei Untersuchungen mit dem Ziel der Prüfung wissenschaftlicher Theorien und Hypothesen eher an die wissenschaftliche Grundlagenforschung ([…]), so ist die *Evaluationsforschung* zweifellos anwendungsbezogen."; zu beiden unten § 2 A.II.2.

[154] *Stanovich*, Psychology 2012, 121: "In [.] applied research, in which the results are intended to be extrapolated directly to a naturalistic situation, questions of the randomness of the sample and the representativeness of the conditions are important".

[155] *Diekmann*, Sozialforschung 2012, 72, und weiter: „Aber auch bei ‚reiner' wissenschaftlicher Grundlagenforschung ohne Auftraggeber können sich die Interessen des Forschers in seinen Befunden bemerkbar machen, und sei es nur das Interesse an der Bestätigung seiner Theorie."

[156] *Diekmann*, Sozialforschung 2012, 38; ähnl. *Baldwin/Davis* in: Cane/Tushnet, Hdb Legal Studies 2003, 880, 897.

[157] Vorbildlich *Bayer/Hoffmann/Sawada*, ZIP 2012, 897 im eigenen Titel („im Auftrag des BMJ") und auf 902 Tabelle 1 unter der Rubrik „Hintergrund der Autoren".

[158] Dazu *Redeker*, NJW 1983, 1034, 1035; *Ulmer*, NJW 1983, 2923, 2924; *Sendler* in: Rüthers/Stern, FG Rechtspolitik 1984, 413, 431 f.; *Habscheid*, NJW 1999, 2230, 2233; vgl. auch *Thieme*, NJW 1983, 2015; *Heinze*, BB 1984, 502.

den: Sie bezweckt die Evaluation[159] etablierter Methoden und gibt praktizierenden Ärzten „Rezeptwissen" an die Hand.[160] Auch im Recht spielt empirische Evaluationsforschung eine wichtige Rolle,[161] wenngleich die Zukunftsperspektiven der empirischen Rechtsforschung maßgeblich auch davon abhängen, dass

„‚Forschung' in der Wissensgesellschaft nicht schlicht Teil von Regierungskunst, also der Legitimation bestimmter Vorhaben dienende Auftragsforschung ist, sondern ‚freie' Reflexion im Vordergrund steht. Gerade für die Forschung zum Recht ist die Gefahr, gezielt zu gutachten anstatt systematisch zu reflektieren, sehr hoch."[162]

E. Produktion und Rezeption von Empirie

Nachdem nun der Bezugsrahmen der vorliegenden Untersuchung abgesteckt ist, bleibt zu klären, welche Rolle Juristen darin zugedacht ist. Vor einigen Jahren identifizierte ein pointierter Fachaufsatz drei mögliche Rollen, in die Juristen im Umgang mit empirischer Forschung schlüpfen können: Außer der Rolle des Ablehners (*critic*), der „aktiv oder passiv, aufgeschlossen oder entschlossen" den „Wert empirischer Forschung in Abrede stellt",[163] sind dies die Rolle des Machers (*doer*), der „versucht, empirische Methoden in der eigenen Forschung einzusetzen"[164] und die des Nutzers (*user*), der „fremde empirische Arbeiten zur Unterfütterung seiner Argumentation zitiert".[165] Eine ähnliche Differenzierung findet sich in der evidenzbasierten Medizin.[166] Dort richtet sich die Evidenzbasierung vorrangig an „Nutzer",[167] wenn auch betont wird, dass sie durch die Aufdeckung von Wissenslücken produktive Forschung anstoßen könne.[168]

---

[159] Vgl. *Timmer* (zit. oben bei Fn. 36); näher unten § 2 A.II.2.

[160] Zum Begriff *Stanovich*, Psychology 2012, 190: "In fact, the idea of recipe knowledge provides one way of conceptualizing the difference between basic and applied research."

[161] Vgl. unten § 1 F.II., insb. bei Fn. 233 und nach Fn. 237, sowie bei Fn. 247.

[162] *Baer* in: Grundmann u.a., FS Juristische Fakultät 2010, 917, 927.

[163] *Diamond*, U Ill L Rev 2002, 803, 808 f.: "The critic may be active or passive, engaging or refusing to engage and debate the value of empirical work."

[164] *Diamond*, U Ill L Rev 2002, 803, 808 f.: "A minority within the legal academy, the doer […] attempts to use empirical methods in conducting her research."

[165] *Diamond*, U Ill L Rev 2002, 803, 809: "the *User* does not conduct empirical research, but cites the empirical work of others to support arguments the user wishes to make."

[166] *Straus/McAlister*, Can Med Assoc J 2000, 837: "doing mode", "using mode" und "replicating mode".

[167] *Del Mar/Glasziou/Mayer*, Brit Med J 2004, 989: "focuses on […] users of research rather than producers of it."; *Straus/McAlister*, Can Med Assoc J 2000, 837, 839: "Although a minority of practitioners of evidence-based medicine also do research, its practice is […] not a method for performing research."

[168] *Rosenberg/Donald*, Brit Med J 1995, 1122, 1125.

Auch für die evidenzbasierte Jurisprudenz ist deshalb zu fragen, inwieweit ihr Erkenntnisinteresse die Übernahme der Rollen von „Macher“ und „Nutzer“, also von Produzent (I.) und Rezipient (II.) empirischer Forschung, erfordert. Im Anschluss ist noch kurz auf eine weitere Rolle einzugehen, der bislang vielleicht weniger Aufmerksamkeit zugekommen ist als sie verdient (III.).

I. Juristen als Produzenten empirischer Forschung

Was ist Rechtswissenschaft, worin liegt ihr *Proprium*? Dazu war vor einigen Jahren zu lesen: „Juristen falsifizieren keine Hypothesen. Juristen nutzen keine Statistiken. Juristen führen keine Interviews. Juristen machen keine Experimente.“ – und doch: „Zu jeder dieser Aussagen gibt es Ausnahmen.“[169] Denn oft „wünschen sich die Juristen [.] verhaltenswissenschaftliche Befunde, die noch niemand erhoben hat“,[170] weil beispielsweise eine empirische Erhebung zwar aus juristischer Sicht, nicht aber aus derjenigen von Sozialforschern lohnenswert erscheint.[171] Bisweilen wird angenommen, dass dies bei juristisch relevanten Fragen „fast immer der Fall sein“ werde, „weil Sozialwissenschaftler für juristische Tagesfragen schwer zu interessieren sind und ihre Problemmoden wohl bisher (noch) in anderen Zyklen verlaufen als die der Jurisprudenz.“[172]

Deshalb können Juristen erwägen, selbst in die Rolle des empirischen Forschers zu schlüpfen:

> „Rechtswissenschaftler, die sich für eine bestimmte Frage interessieren, werden oft feststellen, dass sie selbst eigene Datenerhebungen durchführen müssen, wenn sie ein Problem empirisch beleuchten wollen.“[173]

Diese Feststellung begegnet immer wieder dem Einwand, Juristen fehlte der „komparative Vorteil bei der Durchführung empirischer Studien“, weshalb sie empirische Forschung lieber ausgebildeten Sozialforschern über-

---

[169] *Engel/Schön* in: Engel/Schön, Proprium 2007, IX.

[170] *Engel* in: Engel u.a., Recht und Verhalten 2007, 363, 374; *Monahan/Walker*, Wis L Rev 1991, 569, 570: “The body of existing research on many questions of crucial interest to the law is either nonexistent or [.] plainly inadequate”.

[171] *Schweizer* in: GfK e.V., Jahrbuch 1976, 386, 396; *Schneider/Teitelbaum*, Utah L Rev 2006, 53, 65: “Why not leave empirical research to social scientists who are trained to do it? The short answer is the rule of necessity – social scientists will never do enough of the kinds of research lawmakers and their advisors need.”

[172] *Winter*, Rechtstheorie 1971, 171, 180; *Engel* in: Engel u.a., Recht und Verhalten 2007, 363, 379: „So sind die meisten Psychologen nicht daran interessiert, Steuerungsprobleme zu verstehen“.

[173] *Lempert*, J Emp L Stud 2010, 907, 924; auch nach *Heldrich*, AcP 1986, 74, 76 haben sich die Sozialwissenschaften der Rechtswirklichkeit „nicht so tatkräftig angenommen, daß eigene empirische Untersuchungen von Juristen entbehrlich wären.“

lassen sollten.[174] Andererseits ist diese Kritik sowohl anfechtbar,[175] als auch zweischneidig, weil sie nämlich ebenso gegen den Einsatz ausgebildeter Sozialforscher gewendet werden kann, soweit es um deren Verständnis rechtlicher Probleme geht.[176] Forscher aus empirischen Disziplinen müssen sich bei der Studienplanung „ebenfalls um eine Fülle externer Orientierungen bemühen", daher braucht sich auch der empirisch produktive Jurist nicht grundsätzlich „entgegenhalten zu lassen, er verstehe nichts von den genannten Disziplinen".[177]

Dennoch ist natürlich nicht zu übersehen, dass die empirische „Produktion" durch Juristen institutionell vielfach behindert wird:[178]

Erstens fehlt Juristen die für eine empirisch produktive Betätigung erforderliche *Methodenausbildung*.[179] Als Abhilfe bietet sich vor allem eine Postgraduiertenausbildung oder die gemeinschaftliche Arbeit mit empirisch ausgebildeten Vertretern anderer Disziplinen an. Nicht umsonst wird die zunehmend anzutreffende Mehrautorenschaft als Indiz für das Erstarken des Empirismus in der Rechtswissenschaft gewertet.[180]

---

[174] *Epstein*, Harv L Rev 2002, 1288, 1291: "they are always subject to the same objection: […] lawyers have no comparative advantage in doing empirical work."; *Lempert*, J Emp L Stud 2010, 907, 923: "It will [.] be the rare legal scholar who enjoys a comparative advantage in analyzing [.] data."; *Hopt*, JZ 1975, 341, 343: „Ganz im Gegensatz zum Sozialwissenschaftler wird jedoch der Rechtsanwender nicht im geringsten dazu ausgebildet, Fakten einwandfrei zu ermitteln."

[175] Vgl. *Riegert*, Dick Int L Ann 1983, 1, 54: "The experience lawyers obtain in proving, or even in contemplating the proof of their cases, apparently develops a skill in empirical investigation which equals or exceeds the empirical research skills of many social scientists."; a.A. *Neumann/Krieger*, Clin L Rev 2003, 349, 392: "the urges to argue and to assemble proof from which to argue can prevent us from knowing in a scientific sense."

[176] *Schweizer* in: GfK e.V., Jahrbuch 1976, 386, 393; *Riegert*, Dick Int L Ann 1983, 1, 4: "These researchers are [.] usually less familiar with the actual practice and application of law"; *Neumann/Krieger*, Clin L Rev 2003, 349, 390: "naiveté about law and its institutions"; *Lempert*, J Emp L Stud 2010, 907, 924: "legally-relevant questions that social scientists untrained in the law would miss."

[177] *Lüderssen*, Erfahrung als Rechtsquelle 1972, 17, und allgemeiner: Es sei „allenthalben unumgänglich geworden, sich von ‚ressortspezifischen Verfahrensweisen' zu befreien".

[178] Zum Folgenden ausf. *Schuck*, J L Edu 1989, 323, 331 ff. ("Reasons for the Neglect of Empirical Research"), *Heise*, Pepp L Rev 1999, 807, 815 ff. ("Reasons for This Dearth") und *Neumann/Krieger*, Clin L Rev 2003, 349, 383 ff. ("What Prevents Us From Doing Better?"); vgl. auch *Bryden*, U Colo L Rev 1992, 641, 645 f.; *Rhode*, Harv L Rev 2002, 1327, 1351 ff.

[179] *Schuck*, J L Edu 1989, 323, 333: "empirical work requires more than the analytical skills that are so highly prized on law school faculties."; genauso *Heise*, Pepp L Rev 1999, 807, 817; *Neumann/Krieger*, Clin L Rev 2003, 349, 390: "Without that background, we are apt to make methodological errors with some – but not all – empirical research methods."

[180] So schon der Titel von *Ginsburg/Miles*, U Ill L Rev 2011, 1785: "Empiricism and the Rising Incidence of Coauthorship in Law".

Zweitens bestehen kaum Anreize für Rechtswissenschaftler, sich empirisch zu betätigen, aber starke *Anreize dagegen*. Lehrstühle für empirische Rechtsforschung gibt es nicht, Berufungskommissionen und Fachbeiräte können empirische Arbeit oft weder beurteilen noch wertschätzen, und Vorbilder für erfolgreiche Studien (und deren Präsentation) sind rar. Andererseits bietet gerade dieser Mangel an Vorarbeiten eine Möglichkeit, wirklich originelle wissenschaftliche Beiträge zu leisten.[181]

Drittens erfordert empirische Forschung größere *finanzielle und personelle Ressourcen* als sie den meisten Juristen zur Verfügung stehen:

> „Zu den *Grundbedingungen* erfolgreicher empirischer Arbeit gehören [.] Geld und nochmals Geld, des weiteren geeignetes Forschungspersonal in ausreichender Kapazität und nach Möglichkeit mit fachlichen Mehrfachkompetenzen“.[182]

Empirische Forschung bedarf häufig einer Finanzierung durch Drittmittel[183] und qualifizierter Hilfskräfte (*research assistants*).[184] Das erklärt auch, warum das deutsche Gesellschaftsrecht empirisch bislang vor allem in Form der Rechtstatsachenforschung arbeitet, die sich – obschon so aufwändig, dass sie oft in Dissertationen betrieben wird – noch am ehesten von einer Einzelperson ohne großes Budget bewältigen lässt.[185]

Viertens erfordert empirische Forschung *viel Zeit* und unterliegt *großer Unsicherheit*. Die Planung ist aufwändig und bedarf oft mehrerer Schritte im Zusammenspiel mit der Durchführung (Vorstudie, Hauptstudie, Folgestudien), und sogar die Veröffentlichung dauert oft noch Monate bis Jahre.[186] Für die juristische Praxis wird es indessen selten „vertretbar“ sein, „die Entscheidung von mitunter Jahre dauernden sozialwissenschaftlichen

---

[181] *Heise*, Pepp L Rev 1999, 807, 821 f.: “the relative dearth of empirical legal scholarship generates one critical strategic advantage for young, untenured legal scholars. The underdevelopment in the field creates far greater opportunities for making an original contribution to legal scholarship.”

[182] *Gilles* in: Prütting/Rüssmann, FS Lüke 1997, 139, 144.

[183] *Getman*, J L Edu 1985, 489, 493 zum “demand that you demean yourself seeking funding”; *Bryden*, U Colo L Rev 1992, 641, 645: “dickering with foundation officials”; *Rhode*, Harv L Rev 2002, 1327, 1353: “Few professors have ready access to financial support on the scale necessary for major empirical studies, and fundraising is not a skill that academics generally have an interest in acquiring.”

[184] *Rhode*, Harv L Rev 2002, 1327, 1354; *Schuck*, J L Edu 1989, 323, 331: “Much of empirical research is grunt work. Unlike our colleagues in other disciplines, we cannot easily slough it off on acolytes. We do not have graduate students who must cast about for dissertation topics and whose dependence on us to get them jobs places them abjectly in our debt.”

[185] So das Schlusswort von *Nußbaum*, AcP 1955, 453, 484: „Rechtstatsachenforschung kann sehr wohl von einem Mann betrieben werden, wenn die statistischen Leidenschaften gezügelt werden. Bis heute haben sich auf diesem Felde vereinte Kräfte durchaus noch nicht dem Unternehmen des einzelnen überlegen erwiesen.“

[186] Dazu noch § 2 A.VI.1. a.E. sowie *Bryden*, U Colo L Rev 1992, 641, 645.

Untersuchungen, etwa Beobachtungen, abhängig zu machen".[187] Selbst für weniger akute Rechtsfragen bedeutet die empirische Herangehensweise zumindest große Unsicherheit, denn im Vorhinein lässt sich kaum je abschätzen, wie relevant die gewonnenen Daten sein und ob sie zur Beantwortung der Forschungsfrage beitragen werden.[188] Der konstante Publikationsfluss, den Juristen gewöhnt sind, ist deshalb in der empirischen Forschung kaum durchzuhalten.

Fünftens mag der empirischen Produktion durch Juristen auch die vorherrschende *juristische Denkweise* entgegenstehen. Wissenschaftliche Falsifikation ist Juristen fremd,[189] daher scheuen sie womöglich die Gefahr, veröffentlichte Erkenntnisse im Nachhinein als falsifiziert (und gerade nicht mehr „vertretbar") anerkennen zu müssen.[190] Auch rechtspolitische „Ideologie" mag dabei eine Rolle spielen.[191] Immerhin ist dieses Hindernis nicht im engeren Sinn institutionalisiert, daher genügt insoweit

---

[187] *Hopt*, JZ 1975, 341, 348; *Neumann/Krieger*, Clin L Rev 2003, 349, 394: "In our daily work, we are required to have knowledge that cannot wait for empirical investigation."; Vranken, Jurist's Frame 2006, 45 (Rn. 50): "testable data in civil law is often absent. The lawyer nevertheless reaches a decision – if he is a judge because he is obliged to do so."; *Engel*, JITE 2010, 199, 201: "One cannot wait until a difficult question of basic research is finally answered."; *Bilz*, JITE 2010, 194, 195: "Sometimes policy makers *must* act hastily."; *Goldsmith/Vermeule*, U Chi L Rev 2002, 153, 161: "tradeoff between […] accuracy and rigor […] and timeliness and relevance"; a.A. *Epstein/King*, U Chi L Rev 2002, 191, 207: "The legal profession values timeliness, but no more than biologists racing to be the first to locate a disease-causing gene, physicists trying to outdistance each other in the search for new elementary particles, political scientists forecasting and explaining election results or the behavior of policymakers, or economists explaining a stock market crash."

[188] *Getman*, J L Edu 1985, 489, 493: "the amount of time one needs to invest to do such research is enormous […] There is no way a diligent person of reasonable intelligence can fail to write an impressive law review article. But one may fail totally doing empirical work."; *Neumann/Krieger*, Clin L Rev 2003, 349, 384 f., 386: "With doctrinal research, you can tell early in the research whether the idea or angle you are pursuing will lead to interesting results. But with empirical research, you will not know until near the end what your results will be or whether others might consider the project a worthwhile investment."; *Rhode*, Harv L Rev 2002, 1327, 1354 bei Fn. 142: "Why bother with fieldwork when you can leap instantly into print by reading some prominent cases and commentary in the comfort of your office and saying what you think?"

[189] Krit. zur Idee einer falsifizierenden Rechtsdogmatik *Eidenmüller*, JZ 1999, 53, 58 f. m.w.N.; anders wieder *Müller/Christensen*, Methodik I 2013, 524 (Rn. 479).

[190] *Heise*, Pepp L Rev 1999, 807, 818 f.; *Neumann/Krieger*, Clin L Rev 2003, 349, 391: "Lack of sympathy with the scientific desire to know, even if the results conflict with our preconception".

[191] *Schuck*, J L Edu 1989, 323, 332.

„ein *Umdenken* in der rechtswissenschaftlichen Forschung und Praxis, um eine generelle Akzeptanz und (vermehrte) Einbindung empirischer Ansätze [...] möglich zu machen.“[192]

Diese fünf strukturellen Hindernisse bedeuten nicht zwangsläufig, dass „von Juristen nicht empirische Arbeit derselben Art und Güte wie von Forschern anderer Disziplinen erwartet werden“ dürfte.[193] Selbst methodisch mangelhafte Studien mag man aber mit Blick darauf hinnehmen, dass sie durch allfällige Folgestudien verbessert werden und nichtsdestotrotz neue Fragestellungen erschließen können.[194] Solange Juristen ihre empirische Primärforschung nicht überbewerten und Schlussfolgerungen theoretisch disziplinieren,[195] ist der mögliche Schaden gering. Zudem kann gerade die eigene empirische Betätigung das Methodenverständnis entwickeln helfen,[196] frei nach der medizinischen Faustregel zum Erlernen neuer Operationstechniken: „einmal zusehen, einmal machen, einmal lehren“.[197]

Gleichwohl ist angesichts der aufgezeigten Schwierigkeiten einstweilen davon auszugehen, dass die empirische Primärforschung für Juristen das Mittel dritter Wahl bildet, wenn erstens keine Aussicht darauf besteht, empirisch ausgebildete Sozialwissenschaftler für die Untersuchung rechtlich relevanter Fragen zu gewinnen und zweitens keine anderweitige Forschung existiert, die sich rezipieren ließe. Selbst insoweit lässt sich allerdings ein gewisses Eintauchen in den empirischen Forschungsprozess nicht vermeiden, wenn denn einmal die Frage im Raum steht, *ob* frühere Forschung existiert – und für die jeweilige Rechtsfrage einschlägig ist.

II. Juristen als Rezipienten empirischer Forschung

Wie oben bereits angedeutet (bei Fn. 81) und im gesellschaftsrechtlichen Schrifttum ebenso bekräftigt, ist es eine „genuin rechtswissenschaftliche Aufgabe, Erkenntnisse von Nachbarwissenschaften zu fusionieren und in

---

[192] *Arndt*, Empirie 2008, 85.

[193] So aber *Arndt*, Empirie 2008, 84.

[194] *Hogarth*, J Acc Res 1982, S108, 111: “Experiments help define new questions.”

[195] Dafür dienen die hier noch zu entwickelnden Rezeptionsgrundsätze (§ 2 B.).

[196] *Diamond*, U Ill L Rev 2002, 803, 818: “Some might claim that the only way to really understand science is to do it.”; prägnant schon *Hopt*, JZ 1975, 341, 348: „Die Forderung, dann eben Sachverständige heranzuziehen, ist [.] solange irreal, als der Rechtsanwender nichts darüber weiß, wie die Sozialwissenschaften seine Tätigkeit erleichtern bzw. verbessern könnten. Um eine Frage zu stellen, muß man in gewisser Weise die Antwort schon kennen.“

[197] *Lawless/Robbennolt/Ulen*, Methods 2010, xx bzw. 5 Fn. 12: “watch one; do one; teach one” und weiter 395: “understanding empirical methods is not a spectator’s activity.”; zur EbM Rousseau, Acad Mgmt Rev 2006, 256, 259: “*participation in research increases the salience of the evidence base*. It helps if physicians in the immediate environment have participated in clinical research”.

einen normativen Rahmen zu überführen".[198] Dies setzt freilich ein fundiertes Verständnis von Methoden und Erkenntniswert empirischer Forschung voraus, weshalb „sich Juristen ohne nähere Kenntnis empirischer Methodik nicht die empirische Deutungshoheit über soziale Fakten anmaßen" sollten.[199] Denn der Rezipient empirischer Studien ist genau wie deren Produzent angewiesen auf die „*Entwicklung eines Bewusstseins*, was Empirie bedeutet und wie deren methodische Ansätze einwandfrei zu nutzen sind."[200] Beides gilt mitunter sogar als juristische Grundlagenkompetenz.[201]

Um die Rezeption empirischer Erkenntnisse sinnvoll zu strukturieren, haben die Vertreter der evidenzbasierten Medizin vier Schritte vorgeschlagen,[202] die sich ebensogut für die evidenzbasierte Jurisprudenz eignen:

1. Fragestellung formulieren (*asking*). Dazu muss der juristische Rezipient durch eine gründliche dogmatische Auseinandersetzung diejenigen empirischen Annahmen oder Bezüge herausarbeiten, die der Rechtsfrage an einer der juristischen Schnittstellen (§ 1 F.) zugrundeliegen.[203]
2. Frühere Forschung finden (*accessing*). Dazu muss der Rezipient sozialwissenschaftliche Datenbanken – die sich nicht wesentlich von juristischen wie juris und Beck Online unterscheiden (unten § 2 A.I.2.) – bedienen und mit dem empirischen Berichtsformat umgehen können.
3. Forschung würdigen (*appraising*). Dazu muss der Rezipient Möglichkeiten und Grenzen der herangezogenen Studien reflektieren,[204] ggf. un-

---

[198] *Langenbucher*, ZGR 2012, 314, 315.

[199] *Petersen*, Staat 2010, 435, 451.

[200] *Arndt*, Empirie 2008, 85; *Diamond*, U Ill L Rev 2002, 803, 818: "What is required […] is a deep understanding of legal and scientific reasoning, and an appreciation of what law and science each recognize as strong and weak evidence, and why."

[201] *Starck*, JZ 1972, 609, 614 (These 6): „Zur *Juristenausbildung* gehört die Einübung in die […] empirischen Bezirke der Rechtswissenschaft."; ähnl. *Müller/Christensen*, Methodik I 2013, 528 (Rn. 482); *Petersen*, Staat 2010, 435, 455: „Juristen können [.] nicht gleichzeitig zu exzellenten Norminterpreten und guten Statistikern oder Experimentatoren ausgebildet werden [… doch sollte man] jungen Juristen in der Ausbildung einen kleinen Kompass für empirische Fragestellungen mit auf den Weg zu geben."; *Neumann/Krieger*, Clin L Rev 2003, 349, 395 f.; ausf. *Nard*, Wake Forest L Rev 1995, 347, 365 ff.; *Burns/Hutchinson*, L Teach 2009, 153; *Bradney* in: Cane/Kritzer, Hdb Empirical 2010, 1025 ff.

[202] *Rosenberg/Donald*, Brit Med J 1995, 1122; *Raspe*, ZaeF 1996, 553, 558; *Del Mar/Glasziou/Mayer*, Brit Med J 2004, 989; *Bilger* in: Comberg/Klimm, Allgemeinmedizin 2004, 74, 75.

[203] Mit *Tontrup* in: Engel/Schön, Proprium 2007, 192, 201 „kommt es darauf an, Rechtsfiguren und Institutionen zu identifizieren, in denen das Recht sich [.] auf fehlerhafte mentale Modelle von Wirklichkeit verlässt"; vgl. auch § 3 Fn. 139.

[204] Vgl. *Fleischer* in: Engel/Schön, Proprium 2007, 50, 74: „Gesellschafts- und Kapitalmarktrechtler müssen die Ergebnisse der theoretischen und empirischen Kapitalmarktforschung *interpretieren* und *nicht duplizieren*."

ter Rückgriff auf die noch zu entwickelnden Rezeptionsregeln (§ 2 B.) sowie die Gültigkeitenlehre (§ 3 A.).[205]

4. Feststellungen verarbeiten (*applying*). Zuletzt muss der Rezipient seine Feststellungen zum Stand der empirischen Erkenntnis auf die Rechtsfrage zurückbeziehen. Das ist die „klassische" normative Arbeit, die die Rechtsfolgen zu einem (jetzt besser) bekannten Tatbestand erörtert.

Diese vier Rezeptionsschritte ermöglichen einen Zugang zur empirischen Forschung, der dem Erkenntnisinteresse der Jurisprudenz gerecht wird und dabei weitgehend die ihr eigene Arbeitsweise ermöglicht. Denn er erfolgt „in erster Linie über Texte (vorzugsweise in wissenschaftlichen Fachzeitschriften der Nachbardisziplinen), nicht jedoch über eigene Forschung auf dem angrenzenden Gebiet."[206] Dementsprechend wird eine evidenzbasierte Jurisprudenz (ebenso wie die evidenzbasierte Medizin) hauptsächlich auf die Rezeption empirischer Forschung zurückgreifen.[207]

III. Juristen als Spekulanten empirischer Forschung

Es bedarf keiner weiteren Begründung, dass sich viele Juristen außer Stande sehen werden, eigene empirische Forschung zu betreiben, und dass sich ihre empirischen Fragen auch nicht immer durch eine Rezeption früherer Forschung beantworten lassen. Das bedeutet aber nicht, dass sie unvermittelt in die dritte eingangs eingeführte Rolle des „Ablehners" (bei Fn. 163) verfallen müssten. Vielmehr steht eine vierte Rolle zur Verfügung, die ich als jene des „Spekulanten" kennzeichnen möchte:

> „Unverhohlene richterliche Spekulationen darüber, welche Ergebnisse die empirische Forschung erbringen könnte, haben den Vorteil, Rechtsregeln unter den Vorbehalt künftiger wissenschaftlicher Entwicklungen zu stellen, und solche Entwicklungen sogar anzuregen."[208]

---

[205] *Rosenberg/Donald*, Brit Med J 1995, 1122, 1123: "Mastering critical appraisal entails learning how to ask a few key questions about the validity of the evidence and its relevance […] Its fundamentals can be learnt within a few hours".

[206] *Schön* in: Engel/Schön, Proprium 2007, 313, 315; vgl. auch *Epstein*, Harv L Rev 2002, 1288, 1291: "Lawyers should be able to understand, interpret, and critique the work of social scientists, not replicate it."

[207] *Diamond*, U Ill L Rev 2002, 803, 817: "empirical scholarship in the legal academy will depend more on growth in the population of [.] users than of [.] doers"; ähnl. *Lempert*, J Emp L Stud 2010, 907, 925; *Kaye*, J L Edu 1984, 97, 100: "Because attorneys practice law, not statistics, […] it is more important for them to be critical consumers of statistical arguments than to be statisticians in their own right."; *Lawless/Robbennolt/Ulen*, Methods 2010, 1: "we guess that most of our audience is interested not so much in *doing* empirical work […] as in being able to be *informed and able consumers*".

[208] *Monahan/Walker*, Wis L Rev 1991, 569, 593, mit einem ausgebauten Konzept, wie Juristen „empirische Fragen ohne empirische Antworten" behandeln sollten.

Die Betonung liegt auf „unverhohlen“, denn solche Spekulationen dürfen nicht als „gesicherte Erfahrungssätze“ in die Rechtspraxis zementiert werden, sondern müssen mit offenem Visier ihre Vorläufigkeit anerkennen. Folglich bedarf auch diese Art von Spekulation einer gewissen Evidenzbasierung, denn die Rolle des Spekulanten verlangt letztlich danach, „unter Auswertung vorhandener Ergebnisse Hypothesen aufzustellen und plausibel zu machen“.[209]

Dadurch können Juristen empirische Forschung zu Fragen anregen, die Sozialwissenschaftlern womöglich entgangen sind.[210] Viele vermeintliche Erfahrungssätze der Rechtspraxis ließen sich leichter hinterfragen, „wenn man sie nur explizieren […] würde“.[211] Deshalb spricht viel dafür, interessierte Spekulation als eine weitere legitime Herangehensweise an empirische Forschung zu betrachten, solange die gebotene „Erkenntnisbescheidenheit“[212] gewahrt bleibt (näher noch unten § 2 A.I.1.).

F. Schnittstellen zur Empirie

Wenn Juristen in verschiedenen Rollen an die empirische Forschung herantreten (so eben E.) und die deutsche Rechtswissenschaft gleichzeitig zur Deutung des Rechts als operativ geschlossener Sollensordnung tendiert (so oben B.), fragt sich, welche „Schnittstellen“ zwischen Sein und Sollen zur Verfügung stehen. Diese Frage ist stärker als alle Übrigen an das deutsche Rechtssystem gebunden und stellt sich bspw. im US-amerikanischen Rechtssystem mit seinem ganz anderen Rechtsverständnis (vgl. oben nach Fn. 89) nicht in gleicher Weise.

Die herrschende Rechtsmethodik[213] hält vor allem drei Schnittstellen zur Rezeption empirischer Erkenntnisse bereit: Rechtsdogmatik (I.), Rechtspolitik (II.) und Rechtswissenschaft (III.).[214]

---

[209] *Winter*, Rechtstheorie 1971, 171, 191.

[210] *Eisenberg*, U Ill L Rev 2011, 1713, 1729: “Law professors can [… direct] studies towards aspects of the legal system that are understudied in [.] social science disciplines.”; *Lempert*, J Emp L Stud 2010, 907, 923: “they can see issues ripe for investigation that sociologists, economists, and others will not have on their radar screens for investigation.”

[211] *Hopt*, JZ 1975, 341, 343.

[212] *Jost*, Soziologische Feststellungen 1979, 117; *Heldrich*, AcP 1986, 74, 87.

[213] Anders insb. *Müller/Christensen*, Methodik I 2013, 527 f., zit. oben vor Fn. 100.

[214] *Engel* in: Engel u.a., Recht und Verhalten 2007, 363, 379 ff.; vergleichbare Kategorien bei *Albert* in: Vernunft 2011, 163, 181: „Außer einer an realwissenschaftliche Erkenntnisse anknüpfenden rationalen Rechtskritik [hier III.] wird eine […] sozialtechnologisch orientierte Jurisprudenz dazu beitragen können, die Probleme des Rechts in einer Weise zu behandeln [hier I.], die […] anderen Wissenschaften Rechnung trägt. Auch rechtspolitische Fragen [hier II.] können auf einer solchen Grundlage rationaler Behandlung zugänglich gemacht werden.“

I. Rechtsdogmatik

Dogmatik ist ein Instrument zum Textverständnis, das sich vorrangig am Wortlaut orientiert. Versteht man den „gewöhnlichen Sprachgebrauch" als Grenze der möglichen Auslegung[215] – was zumindest für das Strafrecht nach Art. 103 II GG gilt – stellt sich unweigerlich die Frage, wie einzelne Begriffe „gewöhnlich" verstanden werden. Die Rechtsdogmatik benötigt hier also eine „empirische Sprachgebrauchsanalyse", um überhaupt die methodischen Grenzen der Auslegung auszuloten.[216]

Zudem verwenden Rechtstexte oft gezielt solche „Verweisungen auf die soziale Wirklichkeit",[217] die sich erst durch empirische Erhebungen ausfüllen lassen.[218] Im Wirtschaftsrecht gilt das etwa für „die Ermittlung eines Handelsbrauchs im Sinne von § 346 HGB im Wege der repräsentativen Umfrageforschung"[219] und für das „Marken- und Wettbewerbsrecht, wo die Demoskopie als Beweismittel vor Gericht anerkannt ist".[220] Demoskopische Institute bieten die „Empirische Rechtsforschung" sogar als eigenständiges Produkt für die forensische Praxis an,[221] weil

> „Begriffe wie die Verkehrsgeltung eines Firmenzeichens, Täuschungsgefahr von Beschaffenheits- und geographischen Herkunftsangaben, Alleinstellung einer berühmten Marke, Verkehrsdurchsetzung einer Ausstattung usw. zum täglichen Brot des Rechtsanwenders gehören und häufig erst unter Benutzung von Repräsentativumfragen erfaßt werden."[222]

---

[215] So statt aller BVerfGE 73, 206, 344 f.; *Larenz/Canaris*, Methodenlehre 2008, 141 ff.

[216] Ausf. *Lorenz/Pietzcker/Pietzcker*, NStZ 2005, 429; weitere Nachw. aus dem rechtslinguistischen Schrifttum bei *Hamann* in: Vogel, Rechtssemantik 2014, i.E.

[217] *Heldrich*, AcP 1986, 74, 91 mit dem Bsp. „ortsübliche Vergleichsmiete".

[218] Zahlreiche Bsp. bei *Schweizer* in: GfK e.V., Jahrbuch 1976, 386, 387 f.

[219] *Eidenmüller*, JZ 1999, 53, 57; ausf. *Oestmann*, JZ 2003, 285; vgl. auch *Schweizer* in: GfK e.V., Jahrbuch 1976, 386, 388 f., auch zur „Empirischen Handelsrechtsforschung" i.R.v. § 87 III HGB (ebd. 406, 417); *Hopt*, JZ 1975, 341, 344; *Heldrich*, AcP 1986, 74, 92.

[220] *Benda/Krenzer*, JZ 1972, 497; *Heldrich*, AcP 1986, 74, 88 f. m.w.N. in Fn. 66; *Pflüger* in: Brand/Strempel, FS Blankenburg 1998, 561 f. (mit Bsp. 566 f.); *Lorenz/Pietzcker/Pietzcker*, NStZ 2005, 429, 430 Fn. 18; *Pflüger*, GRUR-Prax 2011, 51 (und frühere Texte derselben Autorin); *Diekmann*, Sozialforschung 2012, 23 f.; *Viken*, NIR 2012, 220.

[221] Am spezialisiertesten wohl Pflüger Rechtsforschung GmbH Institut für Rechtsdemoskopie (www.rechtsforschung.de/institut.html), ehem. TNS Infratest (www.tns-infratest.com/branchen_und_maerkte/rechtsforschung.asp); ebenso GfK Marktforschung (www.gfk.com/marktforschung/services/empirical_law_research), UfU Unabhängiges Institut für Umweltfragen (www.ufu.de/de/forschung/empirische-rechtsforschung.html), Ipsos (www.ipsos.de/default.asp?c=32); IfD Allensbach (www.ifd-allensbach.de/leistungsspektrum/arbeitsbereiche/umfragegutachten-fuer-die-rechtspraxis.html); letzter Zugriff jeweils am 16.2.2014.

[222] *Hopt*, JZ 1975, 341, 344; gleiches gelte für „Begriffe wie Wettbewerb, gesamtwirtschaftliches Gleichgewicht, angemessenes Wirtschaftswachstum", die „sich im Wirtschaftsrecht allenthalben" fänden (ebd. 345).

Auch wo ein Gesetz solche ausfüllungsbedürftigen Begriffe nicht gezielt verwendet, aber implizit „auf die Meinungen von bestimmten Verkehrskreisen oder Bevölkerungsgruppen" abstellt,[223] könnte empirische Forschung Platz greifen, etwa zur Bestimmung der „im Verkehr als wesentlich angesehenen" Eigenschaften einer Sache (§ 119 II BGB) sowie zur Frage, welche Pflichtverletzungen „unerheblich" sind (§§ 281 I 3, 323 V 2 BGB).[224] Auch das Bestehen einer „Verkehrssitte" im Rahmen der Auslegung nach § 157 BGB unterliegt der tatrichterlichen Feststellung[225] und macht genau genommen empirische Forschung „unverzichtbar".[226]

Wo jedoch der Wortlaut nicht unmittelbar auf Empirie verweist, kann sie

> „dogmatisch nur Bedeutung erlangen, weil sie zur Interpretation des Textes beiträgt. Das kann auf zwei Arten geschehen. Entweder hilft [… sie], das Telos der Norm besser zu verstehen oder den Sachverhalt."[227]

Das damit angesprochene Ziel (griech. τέλος *telos*) der Norm ist durch eine Auslegung nach Sinn und Zweck zu ermitteln. Diese Auslegung ist in zwei Spielarten denkbar,[228] die beide empirisch bedingt sind: Die subjektive Spielart orientiert sich ausschließlich an dem Zweck, den der historische Gesetzgeber vor Augen zu haben meinte, und muss letztlich fragen, „welche Auslegung diesen Zweck am besten erfüllt – was in erster Linie eine faktische Untersuchung ist."[229] Die objektive Spielart dagegen, die den Normzweck verabsolutiert und nach dem heutigen Sinn einer Vorschrift fragt, kann diesen Sinn nur ermitteln, wenn sie weiß, auf welches tatsächliche Problem die Norm überhaupt bezogen sein kann und „welche […] tatsächlichen Wirkungen die jeweilige Norm in dem von ihr geregelten Bereich überhaupt begründet."[230] Also benötigt auch sie die

> „Erkenntnismethoden der Sozialwissenschaften, um das Problem zu erkennen, das zu lösen ist. […] Kausale Erklärungen, funktionale und normative Bewertungen fließen ein in

---

[223] *Heldrich*, AcP 1986, 74, 88, sowie 90: „derartige Tatbestandsmerkmale erfordern [.] an sich die Ermittlung der tatsächlichen Vorstellungen der beteiligten Verkehrskreise."; anderes gelte aber (ebd. 93 ff.) für die „guten Sitten" in § 138 BGB und § 1 UWG, die z.B. *Hopt*, JZ 1975, 341, 344 m.w.N. empirisch konkretisieren wollte.

[224] *Heldrich*, AcP 1986, 74, 88 (damals noch bezogen auf § 459 I BGB a.F.).

[225] So ausdr. BGH, LM § 157 BGB Nr. 1 .

[226] *Heldrich*, AcP 1986, 74, 92; *Schweizer*, Wirklichkeit 2000, 51: „So gut wie in jedem Rechtsgebiet wird ständig auf ‚den Verkehr' abgestellt; und es fragt sich eben, ob darauf verzichtet werden kann, die pluralistische Wirklichkeit zu beachten."

[227] *Engel* in: Engel u.a., Recht und Verhalten 2007, 363, 379.

[228] Dazu *Walz*, ZJS 2010, 482, 483 ff.; *Larenz/Canaris*, Methodenlehre 2008, 149 ff.

[229] *Petersen*, Staat 2010, 435, 440; nach *Tontrup* in: Engel/Schön, Proprium 2007, 192, 201 „kann die Sozialwissenschaft auf verdeckte Folgen aufmerksam machen, die dem teleologisch orientierten juristischen Diskurs entgehen."

[230] *Gusy*, JZ 1991, 213.

die Erkenntnis der Sachlage. Daraus formuliert der Jurist sein Verständnis vom realen Problem. Insoweit ist der Jurist Empirist im Sinne Webers.“[231]

Mithin ist die Auslegung nach dem Sinn und Zweck stets empirisch bedingt.[232] Im Gesellschaftsrecht beispielsweise lässt sich fragen, wann die Vergütung des Vorstands einer Aktiengesellschaft nach § 87 I 1 AktG „in einem angemessenen Verhältnis zu den […] Leistungen des Vorstandsmitglieds“ steht. Die Auslegung dieser Vorschrift mit juristischen Bordmitteln kann zwar zu einer näheren Klärung des Angemessenheitsbegriffs führen (dazu § 3 B.I.), aber zentraler Zweck der Norm bleibt das „Verhältnis“ zwischen Vergütung und Leistung – zwei empirischen Größen. Dieses Verhältnis betrachte ich näher unter § 3 B.V.; ein weiteres Beispiel zur rechtsdogmatischen Rezeption empirischer Erkenntnis erörtere ich unter § 4 D.II.

II. Rechtspolitik

Rechtspolitik ist ein Instrument zur Rechtsgestaltung, also zur adäquaten Regelung der tatsächlichen Verhältnisse. Sie eröffnet zwei Einsatzbereiche für die empirische Forschung, nämlich einerseits die „Feststellung der sozialen Wirklichkeit als Regelungshintergrund eines Gesetzentwurfs (Orientierungs- und Informationsforschung)“ und andererseits die „Kontrolle des Vollzuges des erlassenen Gesetzes (Wirkungs- oder Evaluationsforschung)“.[233]

Der als *Orientierungsforschung* bezeichnete Einsatz empirischer Methoden in der Rechtspolitik fragt etwa:

„Gibt es überhaupt ein regelungsbedürftiges Problem? Welches ist die beste Lösung? Werden sich eine normative angemessene Problemdefinition und eine normativ befriedigenden Lösung im politischen Prozess durchsetzen? Wenn zu befürchten ist, dass dies scheitert: lassen sich die politischen Institutionen so verändern, dass der normativ gewünschte Zustand wahrscheinlicher wird?“[234]

All diese Fragen lassen sich überhaupt nur empirisch beantworten. Deshalb gehören zu den drei „wesentlichen Entscheidungsgrundlagen“ für die Aus-

---

[231] *Mastronardi*, Juristisches Denken 2003, Rn. 334; ähnl. schon Rn. 287: „Rechtswissenschaft ist zwar eine normative Wissenschaft. Aber nicht nur: sie muss auch die soziale Wirklichkeit erkennen. Sie ist damit Geisteswissenschaft und Sozialwissenschaft in einem, ein Verfahren der gegenseitigen Übersetzung zwischen Norm und Faktum.“

[232] Vgl. auch *Eidenmüller*, JZ 1999, 53, 58: „Immer dann, wenn *teleologische Erwägungen*, die sich auf die Fernwirkungen von gerichtlichen Entscheidungen beziehen, zum *Regelvollzug* gehören, ist eine realwissenschaftliche Folgenprognose methodisch geboten.“

[233] *Strempel* in: Broda u.a., FS Wassermann 1985, 223, 232; *Heldrich*, AcP 1986, 74, 110; *Zeisel*, JZ 1974, 561: „Zumindest werden gute Daten ein gutes Gesetz immer besser machen.“

[234] *Engel* in: Engel u.a., Recht und Verhalten 2007, 363, 383; ähnl. *Petersen*, Staat 2010, 435, 437: „Man kann nur sinnvoll entscheiden, was sein soll, wenn man weiß, was überhaupt tatsächlich möglich ist.“

arbeitung von Gesetzentwürfen nicht zuletzt die „Ergebnisse empirischer Rechtsforschung (Ressortforschung, Grundlagenforschung)“,[235] „die insoweit auch Teil der Gesetzgebungslehre ist“[236] und durch die verfassungsrechtliche Verhältnismäßigkeitsprüfung wiederum auf die Dogmatik zurückwirkt.[237]

Ebenso verfassungsrechtlich geprägt ist der als *Evaluationsforschung* bezeichnete Einsatz empirischer Methoden, denn die Rechtspolitik muss laufend gewärtigen, dass

> „ein bei Erlaß verfassungsmäßiges Gesetz nachträglich verfassungswidrig wird, weil sich die tatsächlichen Verhältnisse, auf die es einwirkt, grundlegend gewandelt haben oder sich die beim Erlaß des Gesetzes verfassungsrechtlich unbedenkliche Einschätzung seiner künftigen Wirkungen später als ganz oder teilweise falsch erweist.“[238]

Daher bedarf die Rechtspolitik empirischer Erkenntnisse über die „tatsächlichen Verhältnisse“ und wie Gesetze darauf einwirken. Das gilt insbesondere, wenn auch bei weitem nicht nur, im Gesellschaftsrecht.[239] Dort wurde beispielsweise vor dem Hintergrund schleichender Veränderungen des aktienrechtlichen Organisationsgefüges bereits mehrfach hinterfragt, welchen Eigenwert das so genannte Kollegialprinzip für die Aktiengesellschaft hat. Allgemeiner noch lässt sich diese Frage auf alle Verbände ausdehnen, in denen ein solches Prinzip gilt. Es beruht zumeist auf einigen empirischen Annahmen, die ich in § 4 mit der empirischen Forschung abgleichen und unter § 4 D.III. rechtspolitisch hinterfragen werde. Auch das bereits erwähnte Verhältnis zwischen Vorstandsvergütung und -leistung hat eine rechtspolitische Dimension, die ich unter § 3 B.I. darstellen werde.

---

[235] *Strempel* in: Broda u.a., FS Wassermann 1985, 223, 232 und weiter: „An dieser Trias wird deutlich, wie groß die Einflussmöglichkeit empirischer Rechtsforschung auf die Gesetzgebung und damit auf einen wesentlichen Teil der gesamten Rechtspolitik ist.“; ebenso *Strempel* in: Schäffer, Rechtskultur 1987, 87, 92.

[236] *Strempel* in: Schäffer, Rechtskultur 1987, 87; ähnl. schon *Schweizer* in: GfK e.V., Jahrbuch 1976, 386, 404; *Gilles* in: Prütting/Rüssmann, FS Lüke 1997, 139, 145.

[237] *Petersen*, Staat 2010, 435, 442 ff.; vgl. auch *Eidenmüller*, JZ 1999, 53, 54: „Realwissenschaft kann [dem Gesetzgeber] unter Umständen zeigen, dass sich ein bestimmtes Ziel mit dem geplanten Mittel überhaupt nicht oder nur unter Inkaufnahme von unerwünschten Nebeneffekten erreichen lässt.“

[238] BVerfGE 88, 203, 309 f. m.w.N. aus der früheren Rspr.; zur „Nachkontrolle“ durch empirische Rechtsforschung schon *Schweizer* in: GfK e.V., Jahrbuch 1976, 386, 406.

[239] Nach *Fleischer* in: Engel/Schön, Proprium 2007, 50, 71 f. ist „der Ruf nach mehr Empirie [.] selbstverständlich nicht auf aktien- und kapitalmarktrechtliche Maßnahmegesetze beschränkt. Er gilt generell für die Realfolgen gesetzlicher Regelungen und führt uns die ‚sozialtechnologische‘ Dimension der Rechtswissenschaft vor Augen, die keineswegs eine Selbstbeschränkung, sondern eine Erweiterung ihres Potentials bedeutet.“; ebenso *Eidenmüller*, ZGR 2007, 484, 492 f. = *Eidenmüller*, JZ 2007, 487, 490 f.

III. Rechtswissenschaft

Der Begriff der Rechtswissenschaft ist mehrdeutig. Er kann einerseits die Wissenschaft vom geltenden Recht (*de lege lata*) oder vom möglichen Recht (*de lege ferenda*) meinen, dann ist Rechtswissenschaft synonym zur Rechtsdogmatik oder Rechtspolitik im eben erläuterten Sinn. Als Rechtswissenschaft lässt sich aber auch die Wissenschaft von der Funktionsweise des Rechts *an sich* verstehen, als Instrument der sozialen Gestaltung und Steuerung (Metajurisprudenz).[240] Dann blickt der Rechts-Wissenschaftler aus der Außenperspektive auf das Recht, „so wie sich ein Politikwissenschaftler die Politik gewählt hat und ein Verwaltungswissenschaftler die Verwaltung."[241] Empirische Forschung bezieht sich hier entweder auf das Verhalten der Regelungsadressaten oder auf das Verhalten der Regelungsurheber selbst.

Als Beispiel aus dem Gesellschaftsrecht mag die Frage dienen, ob zur Erreichung einer gesellschaftspolitischen Zielvorstellung die gesetzliche Fremdregulierung oder die private Selbstregulierung von Unternehmen wirksamer ist.[242] Diese Frage betrifft die Effektivität von sozialen Steuerungsinstrumenten überhaupt und steht damit als Metafrage über bzw. vor dem tatsächlich geltenden oder möglichen Recht. In der konkreten Untersuchung mag sie natürlich im Gewand einer spezifischen Rechtsnorm erscheinen – etwa in Form des Deutschen Corporate Governance Kodex – und insoweit auch rechtspolitische Brisanz gewinnen, sie hat aber grundsätzlich darüber hinausweisende Bedeutung.

G. Empirie und Gesellschaftsrecht

Nachdem nun die allgemeinen Querbezüge zwischen empirischer Forschung und Recht geklärt sind, soll der letzte Abschnitt dieses Kapitels konkret die empirische Gesellschaftsrechtsforschung untersuchen.

Schon lange ist anerkannt, dass das Gesellschaftsrecht „mehr als andere Rechtsgebiete auf außerrechtlichen, wirtschaftlichen und sozialen Grundlagen" beruht und „daher auch im besonderen Maße der Einwirkung dieser außerrechtlichen Seinselemente ausgesetzt" ist.[243] In neuerer Zeit befassen

---

[240] Dieses Verständnis legt wohl *Diamond*, U Ill L Rev 2002, 803, 33 f. zugrunde, wenn sie als drei von vier „generellen" empirischen Facetten des Rechts erstens rechtspolitische Institutionen, zweitens das Rechtsbewusstsein und drittens die gesellschaftliche Bedingtheit von Rechtsverletzungen aufführt.

[241] *Engel* in: Engel u.a., Recht und Verhalten 2007, 363, 386.

[242] Empirisch untersucht z.B. von *Fischer/Goerg/Hamann*, SSRN 2012.

[243] So schon *Geiler* in: Predari u.a., Beiträge 1927, 593, 598; vgl. auch *Strine*, Del J Corp L 2002, 499 zur "Inescapably Empirical Foundation of the Common Law of Corporations".

sich empirische Disziplinen vor allem unter dem Schlagwort *corporate governance* mit gesellschaftsrechtlichen Fragen.[244] Dementsprechend gehören Gesellschaftsrechtler zu den interessiertesten Rezipienten empirischer Forschung.[245] So formulierte Hanno Merkt als eines von „zwei zentralen Postulaten" zur Zukunft der deutschen Unternehmensrechtsforschung, dass sie einer „verstärkten interdisziplinären Zusammenarbeit" bedürfe, nämlich

„vor allem mit der Empirie und Ökonometrie […] Die aus dieser Kooperation gewonnenen Synergien werden die zukünftige Gestaltung des Rechtsrahmens für Unternehmen und Kapitalmärkte nachhaltig befruchten. Das lässt sich bereits mit den Erkenntnissen exemplifizieren, die der relativ junge Zweig des behavioural finance hervorgebracht hat."[246]

Auch Holger Fleischer zählt mit der Gewinnung von Rechtstatsachen und der Evaluierungsforschung zwei empirische Aufgaben zu den fünf „Zukunftsaufgaben" des Gesellschaftsrechts,[247] denn

„juristische Probleme lassen sich ohne ein klares Bild von der sozialen Wirklichkeit nicht lösen, oft nicht einmal erkennen. Das gilt namentlich für das Gesellschafts- und Kapitalmarktrecht."[248]

Wiewohl deutsche Gesellschaftsrechtler also großes Interesse an der Rezeption empirischer Erkenntnisse haben, fehlt oft das nötige Verständnis der empirischen Methoden und der Überblick über die fremde und sich schnell entwickelnde Forschungslandschaft.[249] Deshalb mag es hilfreich sein, die beiden empirischen Hauptrichtungen auszuleuchten, die im deutschen und im US-amerikanischen Gesellschaftsrecht bisher den größten Einfluss gehabt haben dürften. Die Abgrenzung und Zuordnung einzelner Studien ist nicht immer ganz unzweifelhaft, doch macht die Kategorisierung gewisse Gemeinsamkeiten und systematische Abgrenzungskriterien in der Masse empirischer Arbeiten sichtbar, die auch Methodenunterschiede

---

[244] *Hopt* in: Reimann/Zimmermann, Hdb Comparative Law 2008, 1161, 1163: „voluminous research on corporate governance in economics, sociology, and, most recently, other disciplines such as behavioural sciences and psychology."; vgl. auch ebd. 1184 f.

[245] *Baldwin/Davis* in: Cane/Tushnet, Hdb Legal Studies 2003, 880, 885: "Company law […] is a field in which there are strong interdisciplinary links and in which the fruits of empirical research are routinely employed by academic lawyers"; *Chui* in: McConville/Chui, Methods 2007, 46, 47: "the tradition of quantitative research is strong in […] corporate law".

[246] *Merkt*, ZGR 2007, 532, 541, vgl. auch 534: „Die Suche nach Antworten erfordert daher die interdisziplinäre Kooperation des Unternehmensrechts mit der Ökonomie sowie der Politik-, Sozial- und Verhaltenswissenschaft"; ähnl. *Eidenmüller*, ZGR 2007, 484, 493 = *Eidenmüller*, JZ 2007, 487, 491.

[247] *Fleischer* in: Engel/Schön, Proprium 2007, 50, 74 ff.

[248] *Fleischer* in: Engel/Schön, Proprium 2007, 50, 69 f.

[249] *Baldwin/Davis* in: Cane/Tushnet, Hdb Legal Studies 2003, 880, 885: "empirical investigation tends not to be something that company lawyers carry forward themselves".

und Entwicklungstendenzen verstehen helfen (vgl. schon oben D.II.). Zudem illustriert sie in etwas größerer Tiefe einige mögliche Themen der empirischen Gesellschaftsrechtsforschung und bildet Anschauungsmaterial für die im Anschluss zu entwickelnde Rezeptionslehre.

I. Rechtstatsachenforschung

Die im deutschen Gesellschaftsrecht bislang bedeutendste empirische Richtung ist die Rechtstatsachenforschung. Da der Begriff der „Rechtstatsache" mehrdeutig ist[250] und in manchen Rechtsgebieten „kaum Verbreitung gefunden hat", weshalb er „selbst in Juristenkreisen regelmäßig Nachfragen auslöst",[251] ist zunächst zu klären, was mit Rechtstatsachenforschung gemeint ist.

Als Rechtstatsachenforschung versteht man die „angewandte empirische Rechtssoziologie auf dem Gebiete des gesamten Zivilrechts"[252] – oder kürzer: die Ermittlung der „für das Recht notwendigen Empirie".[253] Aufbauend auf der soziologischen Jurisprudenz[254] und dem Rechtsrealismus wurde sie kurz vor dem ersten Weltkrieg maßgeblich durch den damaligen Rechtsanwalt Arthur Nussbaum begründet[255] und erfüllt für das Zivilrecht die gleiche Funktion wie die Verwaltungslehre für das öffentliche bzw. die Kriminologie für das Strafrecht.[256] Daher ist „Rechtstatsachenforschung" genau genommen keine Bezeichnung für eine bestimmte empirische Richtung, sondern ein allgemeines Synonym für „sämtliche Methoden der Empirischen Sozialforschung" in ihrer Anwendung auf dem Gebiet des Zivilrechts.[257] Auch die Prägung dieser Methoden durch ein soziologisches Vor-

---

[250] Im Zivilprozessrecht werden damit „Rechtsbegriffe (zB Verschulden), -verhältnisse (zB Miete, Erbenstellung) oder -folgen (zB Anspruch, Nichtigkeit)" bezeichnet, soweit sie „in den konkreten Tatsachen selbst oder in deren Subsumtion unter eine Rechtsvorschrift begründet liegen" (*Saenger* in: Saenger, ZPO 2013, § 284 Rn. 11).

[251] *Pflüger* in: Brand/Strempel, FS Blankenburg 1998, 561.

[252] *Röhl*, Rechtssoziologie 1987, 49.

[253] *Chiotellis/Fikentscher* in: Chiotellis/Fikentscher, Rechtstatsachenforschung 1985, 1, 3.

[254] *Chiotellis/Fikentscher* in: Chiotellis/Fikentscher, Rechtstatsachenforschung 1985, 1: „Damals stand die juristische Methodendiskussion unter dem Eindruck von Rudolph v. Iherings Zweckjurisprudenz, Emile Durkheims ‚éléments de sociologie' und Eugen Ehrlichs ‚lebendem Recht'."; *Riegert*, Dick Int L Ann 1983, 1, 13 f. ergänzt noch Max Weber.

[255] *Nussbaum*, Rechtstatsachenforschung 1914; erläutert im Geleitwort von *Rehbinder* in: Nußbaum, Programmschriften 1968, 9 ff.; *Heldrich*, AcP 1986, 74 f. verweist auf eine Pionierarbeit *Martin Wolffs* von 1906.

[256] *Heldrich*, AcP 1986, 74, 77; *Röhl*, Rechtssoziologie 1987, 48; *van Aaken*, Rational Choice 2003, 125 Fn. 541; *Arndt*, Empirie 2008, 36 bei Fn. 181 und *Arndt*, Soft Law 2010, 52 f.; *Lüderssen*, Erfahrung als Rechtsquelle 1972, 14 hingegen versteht die Kriminologie als „Zweig der Rechtstatsachenforschung".

[257] *Pflüger* in: Brand/Strempel, FS Blankenburg 1998, 561, 562.

verständnis ist mithin kein zwingendes Begriffsmerkmal. Zumindest soweit diese Forschung quantitativ betrieben wird, verfügt sie allerdings über so ausgeprägte methodische Grundzüge und kennzeichnende Gemeinsamkeiten,[258] dass sie als abgrenzbare empirische Strömung im weit größeren Kosmos empirischer Methoden aufgefasst werden kann.

Im Gesellschaftsrecht lassen sich nach meiner Deutung fünf Entwicklungsetappen dieser quantitativen Rechtstatsachenforschung ausmachen:[259]

Erste Anfänge. Erste rechtstatsächliche Arbeiten im Gesellschaftsrecht entstanden bereits um die Jahrhundertwende.[260] Der Begründer der Rechtstatsachenforschung Nussbaum selbst war es schließlich, der 1926 die Schriftenreihe „Gesellschaftsrechtliche Abhandlungen" ins Leben rief, die bis 1933 insgesamt dreiundzwanzig „im Sinne der Rechtstatsachenforschung geschriebene" Monographien zusammentrug.[261] Diese Forschung erregte auch international Aufmerksamkeit.[262]

Neubeginn. „Als einer der ersten [nahm] nach dem Kriege Ernst E. Hirsch den Faden empirischer Rechtsforschung in der Bundesrepublik Deutschland wieder auf",[263] indem er im Todesjahr Nussbaums 1964 an dessen Wirkungsstätte Berlin das „Institut für Rechtssoziologie und Rechtstatsachenforschung" gründete.[264] Zwei Jahre später rief er eine Schriftenreihe ins Leben, die seit der darin erfolgten Neuveröffentlichung von Nussbaums Programmschriften[265] als „Schriftenreihe zur Rechtssoziologie und Rechtstatsachenforschung" firmiert und bis heute knapp einhundert Titel verzeichnet. In dieser Reihe erschien als zweite Arbeit, gleich nach dem von Hirsch verfassten Einführungsband, eine gesellschaftsrechtliche Dissertation von Jutta Limbach über „Theorie und Wirklichkeit der GmbH".

---

[258] Es sind meist Beobachtungsstudien mit beschreibendem Erkenntnisinteresse, die auch fast nur beschreibende Statistik verwenden (zu diesen Begriffen näher im nächsten Kapitel); a.A. *Gilles* in: Prütting/Rüssmann, FS Lüke 1997, 139, 147, der keine „jeder Rechtstatsachenforschung gleichermaßen zugrundeliegende Denk- und Arbeitsweise" erkennt.

[259] Studien, die nicht allgemein veröffentlicht wurden, bleiben unberücksichtigt.

[260] Vgl. *Vogel*, Aktienwirklichkeit 1980, 26 ff.; *Lieder* in: Bayer, AG im Spiegel 2007, 79, 80 m.w.N. in Fn. 3.

[261] *Rehbinder* in: Nußbaum, Programmschriften 1968, 9, 14 (Monographien dort Fn. 21).

[262] *Nussbaum*, Colum L Rev 1940, 189; später *Rehbinder*, J L Edu 1972, 567; *Riegert*, Dick Int L Ann 1983, 1 f.: "the most important development in German law in the past two decades […] Social-fact research in law […] is likely to be of increasing importance to lawyers"; für die Türkei *Gilles* in: Prütting/Rüssmann, FS Lüke 1997, 139.

[263] *Strempel* in: Broda u.a., FS Wassermann 1985, 223, 225.

[264] Anschaulich porträtiert bei *Riegert*, Dick Int L Ann 1983, 1, 18 ff., zusammen mit den Rechtstatsacheninstituten von *Dieter Strempel* im Bundesjustizministerium seit 1973, *Volkmar Gessner* am Max-Planck-Institut Hamburg 1970–1982 und *Erhard Blankenburg* am Wissenschaftszentrum Berlin 1969–1980; vgl. auch *Heldrich*, AcP 1986, 74, 111 f.; zu einzelnen Pionierarbeiten noch *Limbach*, JA 1973, 143, 146 ff.

[265] *Nußbaum*, Programmschriften 1968.

Sie wollte „empirische Normaltypen“ jener Rechtsform identifizieren, die am Reißbrett ohne Vorbild entworfen worden war – als ein „Sprung ins Dunkle“ der Rechtswirklichkeit.[266] Mittels eines strukturierten Erfassungsbogens nahm die Autorin Daten zu 486 GmbH auf, die zwischen 1961 und 1963 in das Handelsregister beim Amtsgericht Berlin-Charlottenburg eingetragen worden waren,[267] und stellte fest, dass in der GmbH der Zusammenhang zwischen Herrschaft und Haftung „in der Rechtswirklichkeit vielfach durchbrochen ist“; dieser Durchbrechung habe „der Schöpfer des GmbH-Gesetzes durch eine weitgehende vertragliche Gestaltungsfreiheit selbst den Weg geebnet“, weil er eine falsche Vorstellung vom praktischen Anwendungsbereich der neuen Gesellschaft zugrunde gelegt habe – eine gefährliche „Gleichsetzung von Wertbild und Wirklichkeit“.[268] Genau 40 Jahre nach Limbachs Arbeit erschien eine weitere Untersuchung, die die spätere Fortentwicklung der Tatsachenforschung zur GmbH dokumentiert und aufgrund einer eigenen Fragebogenstudie zum Ergebnis kam, „dass eine gesetzeskonforme Handhabung der GmbH in vielen Fällen offensichtlich misslingt [… und] die Haftungsbeschränkung zumindest gegenüber den Banken nur sehr eingeschränkt“ funktioniere.[269]

Aktienrechtstatsachen. Bald nach der GmbH geriet auch die Aktiengesellschaft in den Fokus der Rechtstatsachenforschung. Das mündete 1980 in ein aktienrechtliches Pendant zu Limbachs Untersuchung der GmbH: Unter dem eindringlichen Titel „Aktienrecht und Aktienwirklichkeit“ ging es der Dissertation von Wolfgang Vogel um die Beurteilung eines weiteren ohne vorherige Erprobung eingeführten Regulierungskonzepts, nämlich der Kompetenzverteilung zwischen Vorstand und Aufsichtsrat in den Aktiengesetzen von 1937 und 1965.[270] Mittels Fragebögen, die an 600 mitbestimmungsfreie Aktiengesellschaften (200 bis 2000 Mitarbeiter) versendet und von etwa der Hälfte beantwortet wurden, untersuchte die Arbeit „zentrale Gegenstände der inneren Verfassung der Aktiengesellschaft“.[271] Sie arbeitete verschiedene Realtypen von Aufsichtsräten und Vorständen heraus, die dem gesetzlichen Idealtyp teils nur bedingt entsprachen.[272] Ein gutes Vier-

[266] *Limbach*, Normaltypen 1966, 12.

[267] *Limbach*, Normaltypen 1966, 32 zitiert nach dem Statistischen Jahrbuch eine Gesamtzahl von 2.691 GmbH in Berlin 1963; ein halbes Jahrhundert später sind es fast zwanzig Mal so viele – *Kornblum*, GmbHR 2012, 728, 694 berichtet 53.085 in 2009, 56.937 in 2010.

[268] *Limbach*, Normaltypen 1966, 109 f.

[269] *Hermes*, Haftungsbeschränkung 2006, 35 ff. (Forschungsbericht), 183 ff., 188 (Fazit) – befragt wurden aber auch nur die Gesellschafter insolventer GmbHs. Zu solchen Auswahleffekten § 3 B.II.2.b).

[270] *Vogel*, Aktienwirklichkeit 1980, 33.

[271] *Vogel*, Aktienwirklichkeit 1980, 7.

[272] *Vogel*, Aktienwirklichkeit 1980, 271 ff.

teljahrhundert später wurden diese und andere Rechtstatsachenerhebungen unter dem Schlagwort der *corporate governance* erneut aufgegriffen.[273]

Systematisierung. Im Jahr nach Vogels Dissertation über die Aktienwirklichkeit hob ein anderes rechtstatsächliches Projekt an: Seit 1981 erhebt der mittlerweile emeritierte Stuttgarter Professor Udo Kornblum so gut wie jährlich Zahlen über Bestand und Verbreitung der deutschen Gesellschaftsformen.[274] Methoden und Umfang der Erhebung haben sich im Lauf der Zeit gewandelt; heute umfasst seine Erhebung die Bestandszahlen aller registerpflichtigen Rechtsformen, die er direkt bei den Landesjustizverwaltungen abfragt.[275] Die so erhobenen Daten leiden zwar unter „ärgerlichen Fehlbuchungen“ und „zum Teil ganz massiven nachträglichen Bestandsberichtigungen“[276] sowie der mehrfachen Zählung von Gesellschaften mit Zweigniederlassungen oder Doppelsitzen,[277] sind aber gegenwärtig die verlässlichste Quelle zur Verbreitung deutscher Handelsgesellschaften und ihrer Bestandsentwicklung im Lauf der Zeit. Schon vor Kornblum hatte Herbert Hansen jahrelang umfangreiche Rechtstatsachen über Aktiengesellschaften[278] und GmbH publiziert,[279] ohne allerdings Quellen und Erhebungsmethoden offenzulegen.[280] In jüngerer Zeit wurde die systematische Erhebung des Rechtsformenbestands vor allem durch einen zweiteiligen Beitrag zur empirischen Forschung über „die GmbH und andere Handelsgesellschaften“,[281] eine Reihe von Beiträgen über die Verbreitung der Societas Europaea,[282] sowie eine Studie über geschlossene Gesellschaften und

---

[273] *Lieder* in: Bayer, AG im Spiegel 2007, 79.

[274] Meist (nicht immer) in der „GmbH-Rundschau“ veröffentlicht – von *Kornblum*, GmbHR 1981, 227 bis *Kornblum*, GmbHR 2013, 693.

[275] *Kornblum*, GmbHR 2012, 728.

[276] *Kornblum*, GmbHR 2012, 728, 732.

[277] *Bayer/Hoffmann*, AG 2009, R30.

[278] Der früheste auffindbare Beitrag war *Hansen*, AG 1971, 148; in der Folge war *Hansen* „ganz lange Zeit als Hauptempiriker der AG tätig“ und hat die heute als „AG-Report“ bekannte Beilage „seinerzeit ganz allein gestellt“, so Dr. Bastian *Schoppe* (Redaktion Die Aktiengesellschaft) vom Verlag Dr. Otto Schmidt KG per Telefon am 10.12.2012.

[279] Der früheste auffindbare Beitrag war *Hansen*, GmbHR 1977, 145, der letzte *Hansen*, GmbHR 2004, 39.

[280] Laut *Schoppe* (Fn. 278) hat *Hansen* nach Auskunft aus dem Redaktionssekretariat seine Quellen „sehr akribisch zusammengetragen“ und stützte sich vor allem auf Geschäftsberichte und Informationen der SdK; womöglich stand er an seinem Wohnort Wiesbaden auch mit dem Statistischen Bundesamt in Kontakt, wo in den „70er Jahren ein Herr Hansen im Bereich Geld und Kredit tätig“ war (Ralf *Müller*, Statistisches Bundesamt, E-Mail vom 7.1.2013).

[281] *Meyer*, GmbHR 2002, 177; *Meyer*, GmbHR 2002, 242

[282] *Eidenmüller/Engert/Hornuf*, AG 2008, 721; *Eidenmüller/Engert/Hornuf*, AG 2009, 845; *Eidenmüller/Engert/Hornuf*, EBOR 2009, 1.

Familienunternehmen fortgesetzt.[283] Auch für das österreichische GmbH-Recht sind zwei rechtstatsächliche Festschriftbeiträge zu verzeichnen.[284]

Institutionalisierung. Einige der größten Fortschritte machte die gesellschaftsrechtliche Tatsachenforschung seit der Gründung des „Instituts für Rechtstatsachenforschung zum Deutschen und Europäischen Unternehmensrecht“ an der Universität Jena 2005.[285] Mit einer umfangreichen Forschungs-, Gutachten- und Veröffentlichungspraxis trugen der Institutsleiter Walter Bayer und seine Mitarbeiter seither zum Fortschritt der rechtstatsächlichen Erkenntnis bei. Das dokumentieren zahlreiche Zeitschriftenbeiträge,[286] mehrere rechtstatsächliche Dissertationen über Themen von Kapitalmaßnahmen über Anleger- und Minderheitenschutz bis hin zur Vorstandsvergütung,[287] sowie der Sammelband „Die Aktiengesellschaft im Spiegel der Rechtstatsachenforschung“.[288]

Neben diesen fünf Entwicklungsphasen (die sich an einer sehr subjektiven Auswahl von Meilensteinen orientieren) regten auch rechtspolitische Entwicklungen stets Rechtstatsachenerhebungen zum Gesellschaftsrecht an. So finden sich in Kirchners „Bibliographie zum Gesellschaftsrecht“ 1965–1991 insgesamt 34 Monographien und 66 Aufsätze, deren Titel oder Autor eine empirische Ausrichtung erkennen lässt (Nachweise im Anhang ab S. 321). Zwar waren darunter nur elf Monographien und acht Aufsätze mit dem Wortstamm „Rechtstats“ im Titel, während die übrigen empirischen Arbeiten zum Teil sogar Nachbardisziplinen angehörten.[289] Dennoch reflektieren alle Arbeiten deutlich die rechtspolitischen Entwicklungen ihrer Zeit: Drei Arbeiten befassten sich mit Ausschließung und Abfindung von Personen- oder GmbH-Gesellschaftern, 14 Arbeiten befassten sich mit Rechnungslegung und Publizität der Kapitalgesellschaften, und ganze 18 Arbeiten widmeten sich der Arbeitnehmermitbestimmung. Auch in jüngerer Zeit ist die Rechtstatsachenforschung eine wichtige Entscheidungsgrundlage für Reformen des geltenden Rechts:

---

[283] *Woywode/Keese/Tänzler*, ZGR 2012, 418.

[284] *Reich-Rohrwig* in: Doralt/Nowotny, FS Kastner 1992, 371; *Reich-Rohrwig* in: Enzinger u.a., FS Frotz 1993, 381.

[285] Historischer Abriss bei *Bayer*, Zahlen 2010, 69 f.

[286] Wohl unvollständige Liste in *Universität Jena*, Publikationen 2014; Gesammelte Beiträge aus der Zeitschrift Die Aktiengesellschaft (AG) bei *Bayer*, Zahlen 2010, 66 ff.

[287] In der Schriftenreihe „Empirische Studien zum deutschen und europäischen Unternehmensrecht“ sind bislang die Dissertationen von *Christoph Bode* (2006), *Matthias Müller* (2006), *Markus Homuth* (2007), *Mathias Hotz* (2007), *Kristian Stange* (2010), *Björn Schigulski* (2010), *Gunnar Dieling* (2011), *Roman Stenzel* (2012) und *Jan Eckert* (2013) erschienen.

[288] *Bayer*, AG im Spiegel 2007.

[289] Bspw. fünf betriebswirtschaftliche Monographien und elf Aufsätze aus den Zeitschriften „Der Betrieb“, „Betriebs-Berater“ und „Schmalenbachs Zeitschrift für betriebswirtschaftliche Forschung“.

Theodor Baums erhob mehrfach Rechtstatsachen zu rechtspolitischen Themen,[290] zuletzt zur aktienrechtlichen Beschlusskontrolle anlässlich seiner Gutachtertätigkeit für den 63. Deutschen Juristentag,[291] die wesentlich zu den Reformgesetzen UMAG und ARUG beitrug.

Wilhelm Niemeier (der schon 1982 mit einer rechtstatsächlichen Arbeit promoviert wurde)[292] beteiligte sich mit „Marktdaten" an der „Reformdebatte" um die Einführung der Ein-Euro-GmbH, die sich inzwischen als „Unternehmergesellschaft" in § 5a GmbHG findet.[293]

Andreas Engert, Lars Hornuf und Horst Eidenmüller untersuchten wiederholt, wie sich der europäische Wettbewerb der Gesellschaftsrechte auf die Anzahl von Neugründungen auswirkt, und plädierten für eine Verringerung der Mindestkapitalanforderungen.[294]

Das Berlin Center of Corporate Governance unter Leitung von Axel von Werder führt im Auftrag der zuständigen Regierungskommission regelmäßige Erhebungen über „Die Akzeptanz der Empfehlungen und Anregungen des Deutschen Corporate Governance Kodex" durch.[295]

Daneben sind verdienstvolle Übersichtsartikel und Monographien zu den Rechtstatsachen der *corporate governance* zu verzeichnen.[296]

II. Ereignisstudienforschung

Wenn in den USA von der „empirischen Erforschung des Gesellschaftsrechts"[297] die Rede ist, dann ist vor allem die Durchführung von Ereignisstudien (*event study*) gemeint. Diese Forschungsmethode entfachte eine regelrechte „Methodenrevolution" in den Wirtschaftswissenschaften,[298] nachdem sie vermutlich 1933 erstmals angewandt und Ende der 1960er Jahre von Finanzwissenschaftlern kanonisiert wurde.[299] Zunächst wurde sie dazu verwendet, die ökonomische Theorie von der Informationseffizienz der

---

[290] *Baums/Fraune*, AG 1995, 97; *Baums* in: Feddersen u.a., Corporate Governance 1996, 324; *Baums/König* in: Forster u.a., FS Kropff 1997, 3; *Baums/Theissen*, ZBB 1999, 125.

[291] *Baums/Vogel/Tacheva*, ZIP 2000, 1649; *Baums/Keinath/Gajek*, ZIP 2007, 1629; *Baums/Drinhausen/Keinath*, ZIP 2011, 2329; krit. zu letzterer *Bayer/Hoffmann/Sawada*, ZIP 2012, 897.

[292] *Niemeier*, Einziehung 1982.

[293] *Niemeier*, ZIP 2006, 2237; *Niemeier*, ZIP 2007, 1794; erster „Zwischenbericht" ebenfalls von *Niemeier* in: Altmeppen u.a., FS Roth 2011, 533; vgl. auch *Eidenmüller*, ZGR 2007, 168, 170 ff.

[294] *Eidenmüller/Engert/Hornuf*, EBOR 2009, 1; *Braun/Eidenmüller u.a.*, ZHR 2013, 131.

[295] Dieser „Kodex Report" erscheint jährlich seit 2003 in „Der Betrieb" und online unter www.bccg.tu-berlin.de/main/publikationen.htm; vgl. auch *Kohl/Rapp/Wolff*, AR 2011, 108.

[296] Insb. *Prigge* in: Hopt u.a., Comparative 1998, 943; *Gerum*, System 2007, passim.

[297] So der Titel von *Bhagat/Romano*, Am L Econ Rev 2002, 380.

[298] *Binder*, Rev Quant Fin Acc 1998, 111: "methodological revolution in accounting and economics as well as finance".

Kapitalmärkte zu überprüfen,[300] doch bald erwies sich auch ihr Wert für die normative Politikfeldforschung (*policy analysis*), insbesondere im Bereich des Gesellschaftsrechts.[301] Mittlerweile existieren Tausende von Ereignisstudien, und für das US-amerikanische Gesellschaftsrecht fasst ein vor zehn Jahren erschienener zweiteiliger Übersichtsartikel ihr Themenspektrum wie folgt zusammen:

„Es gab weit mehr Ereignisstudien über Unternehmensübernahmen als über jedes andere Thema, aber kein wichtiges Thema der Unternehmensführung ist von der Methode unberührt geblieben, bis hin zur grundlegendsten Frage überhaupt, der Genese von Gesellschaftsrecht – genauer gesagt: der Frage, ob der Regelungswettbewerb zwischen den Bundesstaaten von Vorteil ist.“[302]

Im Einzelnen gruppiert der Artikel die gesellschaftsrechtlichen Ereignisstudien unter den Themen Schadensersatzklagen gegen Unternehmen (153 ff.), gesellschaftsrechtlicher Regelungswettbewerb (382 ff.), Unternehmensübernahmen (394 ff.), Wertpapierstreitigkeiten (397 ff.), Vorstandszusammensetzung (401 ff.), Beschlussanträge von Aktionären (404 f.), abgeleitete Aktionärsklagen (405 ff.) sowie Vorstandsvergütung (408 f.).

Die große Bedeutung der Ereignisstudie für das Verständnis und die Gestaltung des US-amerikanischen Gesellschaftsrechts erklärt sich zwanglos aus der „natürlichen Passung“[303] zwischen einer finanzwirtschaftlichen Börsenkursbetrachtung und der Aktionärsnutzenausrichtung (*shareholder value*) des US-amerikanischen Gesellschaftsrechts. Weil „die Gesellschaft“ nach US-amerikanischem Rechtsverständnis (zum deutschen bei Fn. 210) in erster Linie ihren Eigenkapitalgebern dient, muss sich jede Einwirkung auf die Gesellschaftsverfassung (ob durch das Recht oder durch die beteiligten Akteure) daran messen lassen, wie sie sich auf den Anteilswert auswirkt.[304] Der Anteilswert schlägt sich bei börsennotierten Unternehmen im Börsenkurs nieder. Ob also eine Maßnahme rechtspolitisch wünschenswert

---

[299] *MacKinlay*, J Econ Lit 1997, 13 f.; weite Teile der Literatur (vgl. nur *Binder*, Rev Quant Fin Acc 1998, 111) halten *Fama/Fisher u.a.*, Int Econ Rev 1969, 1 für die „Erfinder“ der Methode.

[300] *Bhagat/Romano*, Am L Econ Rev 2002, 141, 142; vgl. bspw. *May*, zfbf 1991, 313.

[301] *MacKinlay*, J Econ Lit 1997, 13, 36: “Perhaps the most successful applications have been in the area of corporate finance.”; *Bhagat/Romano*, Am L Econ Rev 2002, 380, 381: “the corporate area is the policy domain in which event studies have been most widely employed”; mit *Frooman*, Bus Soc 1997, 221 existiert sogar eine Metastudie (dazu unten § 2 A.VII.2.) über 27 Ereignisstudien.

[302] *Bhagat/Romano*, Am L Econ Rev 2002, 141 (144: “Examples of events that have been studied are takeovers, equity offerings, change in state of incorporation, adoption of antitakeover provisions, filing of lawsuits against corporations, deaths of corporate executives, and product recalls.”); *Bhagat/Romano*, Am L Econ Rev 2002, 380 (Zitat auf 382).

[303] *Bhagat/Romano*, Am L Econ Rev 2002, 141, 142; ebenso *Bhagat/Romano*, Am L Econ Rev 2002, 380, 381.

ist, lässt sich nach dieser Lesart dadurch messen, dass man ihre Auswirkungen auf den Börsenkurs ermittelt. Das leistet die Ereignisstudie.

Ihr Grundprinzip ist denkbar einfach:[305] Folgt man der ökonomischen Theorie von der halb- oder mittelstrengen Informationseffizienz dezentral organisierter Kapitalmärkte (*semi-strong efficient market hypothesis*), sollten alle öffentlich bekannten und vorhersehbaren Tatsachen, die den Wert eines Wertpapiers bestimmen, immer bereits eingepreist sein. Deshalb können Kapitalanleger überhaupt nur noch dadurch überzufällige Renditen erzielen, dass sie schnell auf *unvorhergesehene* Ereignisse reagieren. Nur solche Ereignisse können deshalb den Börsenpreis noch messbar erschüttern, wobei sich umgekehrt am Ausmaß der Erschütterung direkt ablesen lässt, wie die Kapitalanleger in ihrer Gesamtheit eine unerwartete Entwicklung beurteilen. Um das Ausmaß der Erschütterung zu bestimmen, geht die Ereignisstudie in vier Schritten vor:[306]

1. Sie definiert ein Ereignis und ermittelt den genauen Zeitpunkt, zu dem die Kapitalmärkte von diesem Ereignis erstmals Kenntnis erlangt haben (*announcement day*).
2. Sie ermittelt die Rendite der betroffenen Aktie nach diesem Zeitpunkt.
3. Sie schätzt anhand historischer Daten, wie hoch die Rendite nach dem Ankündigungstag ausgefallen wäre, wenn das Ereignis an jenem Tag nicht angekündigt worden wäre.
4. Sie ermittelt die Differenz zwischen erwarteter und tatsächlicher Rendite (*abnormal return*[307]) und errechnet, ob diese Differenz statistisch verlässlich (signifikant, dazu noch § 2 A.IV.2.c) ist.

Dieser einfache Vierschritt ist leicht nachvollziehbar, was zur Beliebtheit der Ereignisstudie beigetragen haben dürfte. Das darf aber nicht darüber hinwegtäuschen, dass jeder dieser vier Schritte auf gewaltigen konzeptionellen Annahmen beruht und Schwierigkeiten technischer Art lösen muss, die sich oft als unüberwindbar herausstellen:

Schon der erste Schritt der Ereignisstudie ist in der Praxis keineswegs einfach.[308] Denn der Zeitpunkt, zu dem die Kapitalmärkte erstmals Kennt-

---

[304] *Bhagat/Romano*, Am L Econ Rev 2002, 141, 142: "the benchmark for evaluating the benefit of corporate and securities laws is whether they improve investor welfare"; zurückhaltend *Rachlinski*, Cornell L Rev 2011, 901, 918: "Conflicting themes run through every area of law. Consequently, empirical study alone cannot dictate legal policy."

[305] *Bhagat/Romano*, Am L Econ Rev 2002, 141, 143 ff.

[306] Nach *Bhagat/Romano*, Am L Econ Rev 2002, 141, 143 f.; ebenso *Nowak/Rott/Mahr*, ZGR 2005, 252, 267.

[307] Auf deutsch als „Überrendite" bezeichnet (z.B. *Bayer/Hoffmann/Weinmann*, ZGR 2007, 457, 478), umfasst aber auch Fälle der Unterrendite.

[308] *Bhagat/Romano*, Am L Econ Rev 2002, 141, 144: "Conceptually, the announcement date is straightforward […] However, identification of this date can sometimes be nontrivial."

nis von einem Ereignis erlangen, muss genau festgestellt werden: „Sickern bereits vor dem festgestellten Ereignis Informationen durch, wirkt sich eine spätere offizielle Verlautbarung des Unternehmens schwächer aus.“[309] Einerseits können die scharfen Sanktionen des Insiderrechts nie ganz verhindern, dass kursrelevante Informationen den Kapitalmarkt vorzeitig erreichen, andererseits lässt sich oft ohnehin kaum feststellen, wann das Ereignis erstmals veröffentlicht wurde und ob nicht gleichzeitig andere Kurstreiber bekannt geworden sind.[310] In diesen Fällen bildet die Kursveränderung am Ankündigungstag nicht mehr den reinen „Wert“ des Ereignisses ab. Nicht zuletzt deshalb werden regelmäßig mehrere Unternehmen mit sehr ähnlichen Ereignissen betrachtet,[311] was jedoch erstens eine subjektive Beurteilung erfordert, welche Ereignisse hinreichend ähnlich sind, und zweitens die (strenggenommen unhaltbare) Annahme, dass solcherlei ähnliche Ereignisse im gleichen Markt einander nicht beeinflussen.

Auch der zweite Schritt der Ereignisstudie ist problematisch. Er erfordert eine Annahme darüber, innerhalb welchen Zeitraums ein angekündigtes Ereignis eingepreist wird. Beruhend auf der Effizienzannahme (oben nach Fn. 305) wird häufig nur der nachfolgende Tag betrachtet, was auch der statistischen Auswertung entgegenkommt.[312] Ganz unzweifelhaft ist die Annahme freilich bis heute nicht. Zudem sind viele Informationen bei ihrer ersten Ankündigung weder vollständig noch endgültig, dann wird entweder die Wahl eines längeren Betrachtungszeitraums erforderlich, der das statistische Instrumentarium überfordern kann,[313] oder eine Differenzierung nach unterschiedlichen Zeitpunkten, die neue subjektive Annahmen mit sich bringt.[314]

Im dritten Schritt der Ereignisstudie sind weitere Annahmen erforderlich. Denn zur Berechnung der Rendite, die ohne das Ereignis zu erwarten

---

[309] *Eidenmüller/Engert/Hornuf*, AG 2009, 845, 853.

[310] Das erschwert beispielsweise eine Untersuchung von Hauptversammlungsbeschlüssen mit dieser Methodik, weil die Einladung zur Hauptversammlung immer mehrere Beschlussvorschläge gleichzeitig enthält, § 121 III 2 AktG.

[311] *Bhagat/Romano*, Am L Econ Rev 2002, 141, 146 f. sowie 149: “If a sample of one is considered, it is quite difficult to determine the separate effects on firm value of the announcement and of the unrelated information item(s).”

[312] *Bhagat/Romano*, Am L Econ Rev 2002, 141, 149: “the power of the event study methodology diminishes substantially as the event period is increased from one to just two days.”

[313] *Bhagat/Romano*, Am L Econ Rev 2002, 141, 149 ff. und weiter 164: “it is very difficult to have much confidence in the results of event studies that consider long-horizon returns of several years.”

[314] *Bhagat/Romano*, Am L Econ Rev 2002, 141, 144: “the researcher’s bias and priors on what is a significant or relevant event enter the analysis.”; vgl. auch *Peterson*, Q J Bus Econ 1989, 36, 37.

gewesen wäre, stehen erstens mindestens drei verschiedene Berechnungsmodelle zur Verfügung[315] und bedarf es zweitens einer Festsetzung des Referenzzeitraums, deren Schwierigkeiten aus der Rechtsprechung zur Barabfindung bei Kapitalmaßnahmen hinlänglich bekannt sind.[316]

Der vierte Schritt schließlich ist zwar „nur“ eine statistische Auswertung. Diese ist aber ihrerseits nicht weniger wertungs- und problembehaftet, worauf ich unter § 2 A.IV.2.a) noch eingehen werde.

Damit ist festzuhalten, dass Ereignisstudien eine Vielzahl von Wertungen und sehr individuelle Problemlösungen erfordern.[317] Für deren vertiefte Erörterung ist hier nicht der Platz, daher sei auf die einschlägige Fachliteratur verwiesen.[318]

Verglichen mit der Bedeutung, die Ereignisstudien im US-amerikanischen Gesellschaftsrecht zukommt, ist es durchaus bemerkenswert, dass die Methode für das deutsche Gesellschaftsrecht nahezu überhaupt nicht zu existieren scheint.[319] Juris weist genau *eine* Gerichtsentscheidung aus, die den Begriff der Ereignisstudie oder dessen englisches Äquivalent verwendet – und zwar nur in der ausführlichen Wiedergabe eines Sachverständigengutachtens, dem sich das Gericht in knappen Worten anschloss.[320] Der Online-Katalog der Deutschen Nationalbibliothek verzeichnet zwar einige Dutzend Dissertationen und andere akademische Qualifikationsarbeiten, die Ereignisstudien durchführen, darunter findet sich allerdings keine einzige rechtswissenschaftliche.[321] Lediglich in den gesellschaftsrechtlichen Zeitschriften findet sich eine Hand voll Ereignisstudien – abermals deutlich weniger als in den Nachbardisziplinen[322] – zu den Themen

---

[315] *Peterson*, Q J Bus Econ 1989, 36, 39 ff. (“three classes of techniques […] several variations within each class.”); *Bhagat/Romano*, Am L Econ Rev 2002, 141, 145 f.; *Eidenmüller/Engert/Hornuf*, EBOR 2010, 35, 44; *Nowak/Rott/Mahr*, ZGR 2005, 252, 268 f.; vgl. auch *Bayer/Hoffmann/Weinmann*, Ereignisstudienmethodik 2007.

[316] Zuletzt BGHZ 186, 229 – Stollwerck.

[317] Treffend *Peterson*, Q J Bus Econ 1989, 36: “In performing an event study, researchers face several options at different points in the process. […] Many researchers report that they use standard event study methodology, though there is nothing standard in the published research employing this methodology.” und weiter 57: “In most instances, there is not one correct technique to apply, but many appropriate techniques.”

[318] Z.B. *Holler*, Event-Study-Methodik 2012, 36 ff.; *Jahn*, Ereignisstudien 2007, 24 ff.; *Gerpott/Jakopin*, WiSt 2006, 66; *Kothari/Warner* in: Eckbo, Hdb Corporate Finance 2007, 3; *Binder*, Rev Quant Fin Acc 1998, 111, 112 ff.; *MacKinlay*, J Econ Lit 1997, 13; *McWilliams/Siegel*, Acad Mgmt J 1997, 626; *Peterson*, Q J Bus Econ 1989, 36, insb. 49 ff.

[319] Ausnahmen etwa *Eidenmüller*, ZGR 2007, 484, 494 = *Eidenmüller*, JZ 2007, 487, 492 sowie *Arndt*, Empirie 2008, 60, 77 f. (mit etwas eigenwilligem Begriffsverständnis).

[320] *LG Düsseldorf*, Az. 14 KLs 6/09, juris Rn. 322 f. (insoweit nicht in AG 2011, 722).

[321] Die am 22.2.2014 unter portal.dnb.de durchgeführte Suchanfrage *(ereignisstudie or "event study") and (hsg=04a or hsg=19 or hsg=340)* lieferte keine Treffer.

Entsprechenserklärung: Eine Ereignisstudie von 2005 kam auf Grundlage von 145 Beobachtungen zu dem Ergebnis, dass bei erstmaliger Abgabe der Entsprechenserklärung (§ 161 AktG) im Jahr 2003 weder die Anerkennung des Deutschen Corporate Governance Kodex noch überdurchschnittlich viele Entsprechungen oder Abweichungen zu einer signifikanten Kursbewegung führten – was belege, dass „die für das Enforcement des Kodex angenommene (und erforderliche) Selbstregulierung durch den Kapitalmarkt nicht stattfindet.“[323]

Marktmanipulation: Ein Aufsatz von 2005 schlug vor, durch Ereignisstudien Taterfolg und Kausalität im Rahmen der Marktmanipulation durch fehlerhafte Ad-hoc-Mitteilungen (§ 38 II WpHG) nachzuweisen. Das beruhte auf der Überlegung, „dass die Entscheidung des BGH, den Kurseinwirkungserfolg nicht über eine (praktisch unmögliche) Befragung der Marktteilnehmer nachzuweisen, dergestalt zu verstehen ist, dass bei Veröffentlichung einer kursrelevanten Angabe die Kurseinwirkung grundsätzlich durch eine Ereignisstudie nachzuweisen ist.“[324]

Rückerwerb: Eine Ereignisstudie von 2007 kam auf Grundlage von 104 Beobachtungen zu dem Ergebnis, dass die Ankündigung eines Rückerwerbs eigener Aktien (§ 71 AktG) zwischen 2000 und 2005 einen Kursanstieg um durchschnittlich 4,5 bis 5,4 % verursachte – allerdings seit 2001 mit rückläufigem Trend, den die Autoren darauf zurückführen, „dass sich einige Rückkaufsankündigungen im Nachhinein als unglaubhafte Signale für den Kapitalmarkt herausgestellt haben“.[325]

Segmentwechsel: Eine Ereignisstudie von 2011 kam auf Grundlage von 32 Beobachtungen zu dem Ergebnis, dass ein Segmentwechsel aus dem regulierten Markt in den Freiverkehr (*downgrading*) zwischen 2005 und 2009 einen Kursverlust von durchschnittlich 0,8 % verursachte, „und beweist somit, dass der Kapitalmarkt einen Segmentwechsel – zumindest in einer ersten Reaktion – als negatives Signal einstuft.“[326]

---

[322] Aus der Betriebswirtschaftslehre stammen Ereignisstudien u.a. zu Squeeze-Out (*Wessel/Lahr/Ehrhardt*, M&A Rev 2004, 361) und Börsenabgang (*Eisele/Walter*, zfbf 2006, 337), Rückerwerbsankündigungen (*Schremper*, zfbf 2003, 578) und Ad-hoc-Mitteilungen (*Güttler*, zfbf 2005, 237), Unternehmensübernahmen und Desinvestitionen (*Bartsch/Börner*, zfbf 2007, 2; *Rustige/Grote*, zfbf 2009, 470) sowie Änderungen in der Zusammensetzung der Leitungsorgane (*Grigoleit*, ZPU 2011, 131; *Horsch/Hundt*, Corp Fin 2012, 225). Daneben existieren auch „Ereignisfallstudien“, die nur einen Einzelfall untersuchen (*Merz/Wömpener*, Corp Fin 2012, 268, 269; *Hundt/Horsch*, Corp Fin 2012, 141).

[323] *Nowak/Rott/Mahr*, ZGR 2005, 252, 279; Gegenstimmen bei *Binder*, Regulierungsinstrumente 2011, 223 Fn. 608.

[324] *Hellgardt*, ZIP 2005, 2000, 2006 m.w.N.; vgl. auch *Nowak/Rott/Mahr*, ZGR 2005, 252, 266 Fn. 67.

[325] *Bayer/Hoffmann/Weinmann*, ZGR 2007, 457, 477.

[326] *Beyer/Freystedt*, AG 2011, R80, R82.

Squeeze-Out: Eine Ereignisstudie von 2003 kam auf Grundlage von 16 Beobachtungen zu dem Ergebnis, dass in den zehn Monaten nach Einführung der §§ 327a ff. AktG die öffentliche Ankündigung[327] eines mit mindestens fünf Mio. Euro Barabfindung und mindestens 3 % Prämie auf den letzten Schlusskurs verbundenen Ausschlusses von Minderheitsaktionären (§ 327a I 1 AktG) dazu führte, dass *andere* aktiv börsengehandelte Aktiengesellschaften mit ebenfalls mindestens 95-prozentigem Mehrheitsaktionär einen Kursgewinn von durchschnittlich 0,81 bis 0,95 % in einem Zwei- bis Drei-Tages-Fenster verzeichneten.[328]

Umwandlung: Eine Ereignisstudie von 2009 kam auf Grundlage von 38 Beobachtungen zu dem Ergebnis, dass die Ankündigung einer Umgründung in die Rechtsform der europäischen Aktiengesellschaft *Societas Europaea* (Art. 2 SEVO i.V.m. §§ 5 ff. SEAG) zwischen 2003 und 2008 keine signifikanten Kursbewegungen verursachte, daher ließen die Daten „keinen verlässlichen Schluss zu, ob die neue Europäische Gesellschaft nicht nur Unternehmen und Unternehmensleiter anspricht (soviel wissen wir) sondern auch diversifizierte Aktionäre."[329]

Diese in der deutschen Gesellschaftsrechtsliteratur publizierten Ereignisstudien illustrieren einerseits die bereits erwähnten methodischen Herausforderungen.[330] Andererseits verdeutlichen sie, warum die Ereignisstudienmethodik im deutschen Recht bei weitem nicht die Anerkennung genießt wie in den USA: Während dort die Aktionärszentrierung des Aktienrechts (*shareholder value*) als gegeben hingenommen und nur noch über Fragen der empirischen Methodik verhandelt wird,[331] herrscht hierzulande eine Bezugsgruppenlehre vor (*stakeholder value*), die außer dem Interesse der Aktionäre auch die Interessen anderer Bezugsgruppen berücksichtigen

---

[327] Gemeint ist wohl die Bekanntmachung nach § 327c I AktG, denn laut *Helmis*, ZBB 2003, 161, 169 Fn. 40 wurden nur Ankündigungen „unter Nennung der Höhe der Barabfindung" berücksichtigt, ungeachtet der kurstreibenden Wirkung früherer Absichtserklärungen.

[328] *Helmis*, ZBB 2003, 161, 173 f.

[329] *Eidenmüller/Engert/Hornuf*, EBOR 2010, 35, 48 – mit dem Hinweis, eine frühere Fassung habe noch signifikante Kursgewinne festgestellt: "We attribute this to the greater problems in identifying the correct event day for reincorporations of less prominent firms."; ebenso die deutsche Kurzfassung *Eidenmüller/Engert/Hornuf*, AG 2009, 845, 851 ff., 853.

[330] Etwa hinsichtlich der Länge des Betrachtungszeitraums (dazu oben bei Fn. 312): *Bayer/Hoffmann/Weinmann*, ZGR 2007, 457 berichten Ergebnisse für den Ankündigungstag und ein Ein-Tages-Zeitfenster, *Beyer/Freystedt*, AG 2011, R80, R82 entscheiden sich für ein Drei-Tages-Zeitfenster, und *Eidenmüller/Engert/Hornuf*, EBOR 2010, 35, 46 experimentieren mit 15 Zeitfenstern bis zu fünf Tagen vor und nach dem Ankündigungstag.

[331] *Bhagat/Romano*, Am L Econ Rev 2002, 380, 382: "all sides hold the same normative conception of corporate law, shareholder wealth maximization. […] data can inform debates that are over the means of implementing public policy, as opposed to debates over the ends of public policy." Ähnlich erklärt sich wohl die Beliebtheit der Methode innerhalb der deutschen Betriebswirtschaftslehre (Fn. 322).

will. Diese Interessen sind aber nicht so einfach feststellbar wie der Börsenkurs, deshalb „lässt sich empirisch kaum sinnvoll beurteilen“, ob ein bestimmtes Ereignis „den Unternehmenswert in der Summe für alle stakeholder eher erhöht oder verringert“.[332] Demnach müssen Ereignisstudien im deutschen Gesellschaftsrecht immer in dem Bewusstsein auftreten, nur eines von vielen relevanten Interessen zu berücksichtigen – statt das ideale rechtspolitische Instrument wie in den USA darzustellen, dient die Ereignisstudie hierzulande allenfalls „bescheideneren Erkenntniszielen.“[333]

---

[332] *Eidenmüller/Engert/Hornuf*, AG 2009, 845, 851.
[333] *Eidenmüller/Engert/Hornuf*, AG 2009, 845, 851.

§ 2

Grundlegung einer pragmatischen Rezeptionslehre

„Empirie in der Rechtsdogmatik gibt es [.] nur im mehr oder weniger offenen Deutungs- und Bezugsrahmen der rechtlichen Normen.“[1]
– Christian Starck 1972

„Unabhängig von einem noch hier und da geführten Prinzipienstreit beginnt sich [.] in der Jurisprudenz die Einsicht durchzusetzen, dass überall da, wo der Jurist sich auf »Erfahrung« verlässt, die Erfahrungswissenschaften einen Beitrag [...] leisten können.“[2]
– Maximilian Herberger und Dieter Simon 1980

Definitionsmerkmal der empirischen Forschung ist ihre systematische Herangehensweise an Gegenstände der Lebenswelt.[3] Als eine von zwei Säulen der evidenzbasierten Jurisprudenz (§ 1 A.II.) unterscheidet sie sich also von der anderen – der juristischen Lebenserfahrung – dadurch, dass sie den Erkenntnisprozess formalisiert und dadurch intersubjektiv nachvollziehbar macht. Wer empirische Forschung würdigen und ihre Ergebnisse rezipieren will, muss diesen formalisierten Prozess verstehen und darf ihn nicht „gewissermaßen als *black box*“ begreifen,

„deren Ergebnisse von Juristen nutzbar gemacht werden, ohne die genauen Voraussetzungen der Studie zu kennen, so wie man einen Computer als Arbeitsmittel benutzt, ohne sich mit den technischen Einzelheiten auskennen zu müssen.“[4]

---

[1] *Starck*, JZ 1972, 609, 613 = *Starck*, Verfassungsstaat 1995, 110; ähnl. *Schweizer* in: GfK e.V., Jahrbuch 1976, 386, 401: „Welche Daten benötigt werden, bestimmt die (rechtsmethodische) Auslegung der vom Gesetz verwendeten Begriffe. Die Datenermittlung setzt die Rechtsauslegung voraus“.

[2] *Herberger/Simon*, Wissenschaftstheorie 1980, 373.

[3] Empirische Forschung wird verstanden als “all techniques for systematically gathering, describing, and critically analyzing data (objective information about the world).” (*Lawless/Robbennolt/Ulen*, Methods 2010, 7), “any attempt to gather information in some disciplined way about how the world actually works.” (*Schneider/Teitelbaum*, Utah L Rev 2006, 53, 61) oder “the collection and analysis of data using standard social science methodologies.” (*Neumann/Krieger*, Clin L Rev 2003, 349, 353).

[4] *Petersen*, Staat 2010, 435, 447; dagegen hält es *Schön* in: Engel/Schön, Proprium 2007, 313, 314 für „kaum denkbar, in der Person eines einzelnen Wissenschaftlers neben der für sachkundiges rechtliches Arbeiten erforderlichen Verfügbarkeit der juristischen Quellen zu-

Die allgemeine Erläuterung empirischer Methoden für ein juristisches Publikum wurde meines Wissens bislang nur dreimal monographisch versucht, nämlich in einer deutschen Diplomarbeit unter dem Titel „Empirie in den Rechtswissenschaften – Fluch oder Segen?“[5], im britischen “Oxford Handbook of Empirical Legal Research”[6] sowie im US-amerikanischen Lehrbuch “Empirical Methods in Law”[7].[8] Wiewohl damit das Potential zu lehrbuchhafter Aufbereitung längst nicht ausgeschöpft sein dürfte, ist das nicht Kernanliegen der vorliegenden Arbeit. Vielmehr dient die folgende Darstellung des empirischen Forschungsablaufs (A.) vor allem dazu, Gesichtspunkte herauszuarbeiten, die für eine pragmatische Rezeptionslehre bedeutsam sind. Um das zu erreichen, ohne beim juristischen Leser Vorkenntnisse voraussetzen zu müssen, werde ich drei didaktische Schwerpunkte setzen:

Erstens erörtere ich empirische Methoden nicht im Bezugsrahmen einer einzelnen Fachdisziplin – wie in Lehrbüchern üblich – sondern begreife sie als disziplinenübergreifende *Methode* (vgl. § 1 C.) und erarbeite einen allgemeinen *Idealtyp* quantitativ-empirischer Forschung, der die Einheit der (mindestens Sozial-)Wissenschaften betont und die Transferleistung minimiert, die Juristen für deren Rezeption erbringen müssen.

Zweitens systematisiere ich etablierte Konzepte und Begriffe neu: Ich gliedere den empirischen Forschungsablauf in ein Sieben-Phasen-Modell (A.), das vor allem zwei Phasen betont (die beiden letzten), die in diesem Zusammenhang meist übergangen werden. Zudem lege ich dar, dass die fast immer zusammenhanglos präsentierten „Vor- und Nachteile“ verschiedener empirischen Studienarten besser zu begreifen sind, wenn die Studienarten als Elemente eines Spektrums aufgefasst werden (§ 3 B.IV.).

Drittens entwickle ich eine auf Juristen ausgerichtete Didaktik, indem ich fachsprachliche Formalisierungen vermeide, möglichst viele Parallelen zur juristischen Arbeitsmethode aufzeige und durchweg Beispiele mit (gesellschafts)rechtlichem Bezug anbringe.

Neben diesen didaktischen Eigenleistungen gehe ich über eine lehrbuchhafte Darstellung dadurch hinaus, dass ich den Forschungsablauf nicht nur beschreibe, sondern auch kritisch würdige. Ich werde pragmatische (weil nicht philosophisch fundierte) Rezeptionsgrundsätze zur Würdigung empi-

---

gleich die entsprechende Sicherheit im Umgang mit den in den Nachbardisziplinen relevanten Fakten zu gewährleisten.“

[5] *Arndt*, Empirie 2008.

[6] *Cane/Kritzer*, Hdb Empirical 2010; weitere Handbuchbeiträge bspw. *Baldwin/Davis* in: Cane/Tushnet, Hdb Legal Studies 2003, 880; *Chui* in: McConville/Chui, Methods 2007, 46.

[7] *Lawless/Robbennolt/Ulen*, Methods 2010, dazu unten bei Fn. 353.

[8] *Zeisel/Kaye*, Figures 1997 wollen laut ihrem Untertitel zwar in „empirische Methoden in Recht und Prozess“ einführen, beschränken sich aber weitgehend auf Statistik.

rischer Erkenntnisse entwickeln (B.) und die andernorts entwickelte Lehre der empirischen Gültigkeiten für den juristischen Gebrauch aufbereiten (§ 3 A.). Daraus treten sowohl die Prinzipien als auch die Praktiken empirischer Forschung deutlicher hervor als in vielen Lehrbüchern.

Die im folgenden Abschnitt (A.) dargestellten Erkenntnissen stammen zu wesentlichen Teilen nicht von mir, sondern geben den Stand der Forschungsmethodik wieder. Um die Darstellung nicht mit Nachweisen zu überfrachten, vermeide ich deshalb Fußnoten, wo sie nicht unmittelbar zum Verständnis beitragen oder ausnahmsweise zum Beleg einer ungewöhnlichen Behauptung erforderlich sind. Meine Quellen und die herangezogene Lehrliteratur erläutere ich gebündelt zum Abschluss meiner Darstellung (C.).

A. Die Phasen quantitativ-empirischer Forschung

Wie eingangs erwähnt, gliedere ich den empirischen Forschungsprozess in sieben Phasen. Andere Einteilungen sind natürlich denkbar,[9] wobei es sich ohnehin nur um ein theoretisches Ideal handelt, denn „eine rein schematische Betrachtung des Forschungsablaufs wird der Komplexität des Forschungsprozesses in der Praxis freilich nicht gerecht."[10] So finden sich in der Praxis durchaus Vorgehensweisen, die vom vorliegenden Schema abweichen – mitunter aber gerade deshalb Probleme bereiten.[11]

I. Recherche

Empirische Forschung baut stets schrittweise auf dem jeweils erreichten Erkenntnisstand auf – der bisherigen „Evidenzbasis". Diese Evidenzbasis hat, wie eingangs (§ 1 A.) dargestellt, zwei Säulen: *Lebenserfahrung* einerseits und *empirische Erkenntnis* andererseits. Ich verwende beide Begriffe in einer technischen Bedeutung: Lebenserfahrung bezeichnet die Gesamt-

---

[9] Z.B. *Lawless/Robbennolt/Ulen*, Methods 2010, 367 und *Epstein/Martin* in: Cane/Kritzer, Hdb Empirical 2010, 901, 923 mit vier Phasen; *Atteslander*, Methoden 2010, 17 mit fünf Phasen, *Diekmann*, Sozialforschung 2012, 187, 192 f. mit fünf „groben" Phasen und neun „Schritten", oder *Lercher* in: Barta u.a., Rechtstatsachenforschung 2009, 205 mit neun Phasen; *Chui* in: McConville/Chui, Methods 2007, 46, 52 zählt zwar auch sieben „Schritte", aber andere als hier (und unterteilt in die beiden „Phasen" der Planung und Ausführung).

[10] *Hermes*, Haftungsbeschränkung 2006, 32 m.Verw. in Fn. 56 auf *Atteslander*, Methoden 2010, 41; ebenso *Arndt*, Empirie 2008, 12: der „typische Prozessablauf nur als Orientierung"; auch nach *Diekmann*, Sozialforschung 2012, 197 „werden die einzelnen Phasen nicht immer linear durchlaufen."

[11] Etwa werden Hypothesen (§ 2 A.II.6.) oft nach der Datenerhebung (§ 2 A.III.) angepasst, dazu unter § 2 B.III..

heit der sog. anekdotischen Evidenz, empirische Erkenntnis die Gesamtheit der sog. kumulativen Evidenz.

Anekdotische Evidenz (von griech. ἀν *an* und ἔκδοτος *ekdotos*, nicht veröffentlicht)[12] bezeichnet Daten, die ohne intersubjektiv nachvollziehbares Auswahlkriterium gewonnen werden,[13] also durch Vorlieben, Gewohnheiten und Fähigkeiten desjenigen vorgeprägt sind, der sie erhält.

Kumulative Evidenz (von lat. *cumulare*, anhäufen)[14] bezeichnet Daten, die systematisch und in einem veröffentlichten Verfahren erhoben wurden.[15] Kumulative Evidenz wird dadurch der wissenschaftlichen Bewährung zugänglich, also replizierbar (von lat. *replicare*, wiederholen).[16]

Während Lebenserfahrung für die Rechtswissenschaft oft den Endpunkt empirischer Bemühungen markiert (dazu gleich 1.), steht sie für die empirische Sozialforschung oft am Anfang, indem sie eine Suche nach empirischer Erkenntnis (dazu 2.) überhaupt erst anregt.

1. Anekdotische Evidenz und Lebenserfahrung

Aufgrund der subjektiven Prägung anekdotischer Evidenz beruht Lebenserfahrung stets auf „mehr oder minder plausiblen, aber zumeist intuitiven Annahmen über bestimmte Wirkungszusammenhänge".[17] Dabei gibt es bei der Wahrheitssuche „kein unzuverlässigeres Instrument als die intuitive Alltagsforschung".[18] Denn die Plausibilität, auf die sie sich stützt, ist

> „kein Wahrheitskriterium. Vielmehr ist solches Alltagswissen, sind solche ‚Bauernregeln' oftmals schlichtweg falsch oder nur in bestimmten Situationen anwendbar."[19]

---

[12] Dudenredaktion, Fremdwörterbuch 2000, 99 („Anekdote"); *Harper*, Etymology 2014, Stichwort "anecdote".

[13] *Diamond*, U Ill L Rev 2002, 803, 805: "a weak form of empirical evidence because it typically is presented without information about how the particular instance described in the anecdote was selected, how accurately it is being described, and how representative it is"; strenger *Neumann/Krieger*, Clin L Rev 2003, 349, 380: "anecdotal evidence is not evidence at all because anecdotes have not been subjected to a methodology that tests them for accuracy and representativeness."; vgl. schon *Schweizer* in: GfK e.V., Jahrbuch 1976, 386, 421.

[14] *Harper*, Etymology 2014, Stichwort "cumulative", zum urindogermanischen Ursprung vgl. ebd. "cumulus".

[15] *Garrn*, Rationalität 1986, 113: „intersubjektive Überprüfbarkeit […] der jeweiligen sozialempirischen Annahmen".

[16] Dudenredaktion, Fremdwörterbuch 2000, 1157 („Replikation" und „Replik"); *Harper*, Etymology 2014, Stichworte "replicate" und "reply".

[17] *Fleischer* in: Engel/Schön, Proprium 2007, 50, 71; vgl. auch *Strempel* in: Broda u.a., FS Wassermann 1985, 223, 224 zu „Alltagserfahrungen der jeweiligen Entscheidungsträger".

[18] *Lercher* in: Barta u.a., Rechtstatsachenforschung 2009, 205, 238 m.w.N.

[19] *Arndt*, Empirie 2008, 22; *Lawless/Robbennolt/Ulen*, Methods 2010, 400: "'plausible' and 'true' are not the same thing."; *Heise*, Pepp L Rev 1999, 807, 808: "scholars possess few, if any, mechanisms to assess anecdotal evidence for truthfulness, typicality, or frequency."

Die Fehlbarkeit dieses alltäglichen „Wissens“ wurde in der wissenschaftlichen Literatur oft und bisweilen amüsant demonstriert[20] und durch eine monographische Untersuchung der zivilrichterlichen Entscheidungsfindung untermauert:

> „Die lediglich nach Plausibilitäts- und Evidenzgrundsätzen unsystematisch, weitgehend ohne Reflexion und Diskussion der möglichen Fehlerquellen durchgeführte Überprüfung führt dazu, dass der empirische Wert von Alltagswissensbeständen nicht selten zweifelhaft ist.“[21]

Denn obwohl „in gewissem Sinn […] jeder Mensch Zugang auch zur sozialen Realität“ hat, nehmen „Richter bisweilen recht ungeniert ihre persönliche Sachkunde für die Beurteilung sehr problematischer gesellschaftlicher Zusammenhänge in Anspruch“.[22] Eine solche Beurteilung kann zwar wertvolle Inspirationen für die empirische Forschung bedeuten, muss dazu aber explizit und mit dem offenen Visier der bloßen Spekulation erfolgen (oben § 1 E.III.). Problematisch wird sie hingegen, wo anekdotische Evidenz unausgesprochen und unreflektiert die Grenzen der Tatsachenermittlung bildet, weil die Möglichkeit „einer erfahrungswissenschaftlichen [d.h. empirischen] Überprüfung“ dieses „Alltagswissens“ und seiner „sozialempirischen Annahmen“ ausgeblendet wird.[23]

So begnügt sich der Gesetzgeber nicht selten mit dem Rückgriff auf anekdotische Evidenz, selbst wenn er externe Stellungnahmen einholt: Die angehörten Sachverständigen berichten oft eigene Beobachtungen, für deren Übertragbarkeit nicht mehr Gewähr besteht als das Vertrauen auf die Unvoreingenommenheit und das Einschätzungsvermögen des Sachverständigen. Zwar fließen durch ein solches Verfahren der Datenerhebung durchaus praktische Erfahrungen und Beobachtungen der Sachverständigen in

---

[20] *Saks*, Mich L Rev 1979, 892, 895 f.; *Diekmann*, Sozialforschung 2012, 25 ff. m.Verw. auf *Lazarsfeld*, Publ Opin Q 1949, 377, 379 f.; *Vranken*, Jurist’s Frame 2006, 39 (Rn. 42); weitere Illustration unten § 4 vor Fn. 391.

[21] *Bürkle*, Alltagstheorien 1984, 172; *Davis*, Org Behav Hum Dec Proc 1992, 3: “Intuitions about human behavior generated from personal experience or conventional wisdom are not always wrong, but it is well known that predictions generated from these sources have sometimes fared poorly when confronted with behavioral research results.”

[22] *Heldrich*, AcP 1986, 74, 81, 85; krit. auch *Schweizer* in: GfK e.V., Jahrbuch 1976, 386, 420 f.: „was schon die Spatzen von den Dächern pfeifen: wenn dem Juristen kein Argument einfällt, beruft er sich auf die Lebenserfahrung“; auch *Arndt*, Empirie 2008, 43 warnt vor „eher zufällig ausgewählten (und oft nicht einmal selbst recherchierten) Beispielfällen“, die „oft rein intuitiv und umgangssprachlich (falsch)“ als „empirisch“ gelten.

[23] *Garrn*, Rationalität 1986, 111; *Schneider/Teitelbaum*, Utah L Rev 2006, 53, 68: “Empirical data compete with less reliable but more alluring evidence.”

die Rechtssetzung ein,[24] dadurch schwindet allerdings nicht der Nutzen systematischer empirischer Studien:

„Denn bei der Gesetzesvorbereitung macht es schon einen Unterschied, ob sich die Feststellung der sozialen Wirklichkeit allein an Stellungnahmen aus der Praxis orientiert oder auf eigener Lebenserfahrung beruht oder ob umfassende oder zumindest repräsentative Erkenntnisse über den zu regelnden Bereich sozialen Lebens vorliegen. Im letzten Fall kann Gesetzgebung auf präziserer und realistischerer Basis aufbauen.“[25]

Gerade weil menschliches Einschätzungsvermögen (auch das von Experten) immer fehlbar[26] und die Realität nie so „offensichtlich“ ist, wie sie scheint,[27] sollte sich eine „rechtspolitische Reformagenda […] nicht auf anekdotische, sondern zumindest auf kumulative Evidenz stützen.“[28]

Als illustratives Beispiel – sprich: als anekdotische Evidenz – für die Verwertung anekdotischer und kumulativer Evidenz in der gesellschaftsrechtlichen Gesetzgebung mag etwa das Gesetz zur Unternehmensintegrität und Modernisierung des Anfechtungsrechts (UMAG) dienen.[29] Dieses Gesetz bediente primär die Bedürfnisse der Aktienrechtspraxis und beruhte auf einer eminenten Menge gesammelter Evidenz – zumeist anekdotischer statt kumulativer Art, was an Begründungen wie „in der Praxis sehr unwahrscheinlich“, „gängige Praxis“ oder „Es ist auch nicht bekannt geworden, dass…“ zu erkennen ist.[30]

Etwa wurde durch Art. 1 Nr. 5 UMAG das Hinterlegungserfordernis in § 123 II, III AktG mit der Begründung aufgehoben, in der Praxis würden „Aktien ohnedies nicht hinterlegt, sondern es erfolgt lediglich eine Anmeldung, verbunden mit der Vorlage einer Bankbescheinigung“, die neue Vorschrift vollziehe demnach „nur die geltende Praxis gesetzlich nach“.[31] Statt

---

[24] Vgl. *Zeisel*, JZ 1974, 561: „Gute Gesetze werden in der Regel sorgfältig vorbereitet, zuweilen in Enqueten oder Kommissionen, die systematischen empirischen Untersuchungen sehr nahekommen.“; auch *Schön* in: Engel/Schön, Proprium 2007, 313, 320 meint, dass „jahre- oder jahrzehntelange Erfahrung mit praktischen Sachverhalten […] mit den Ergebnissen manch methodisch härterer empirischen [sic] Forschung mithalten kann.“

[25] *Strempel* in: Schäffer, Rechtskultur 1987, 87, 92; *Strempel* in: Broda u.a., FS Wassermann 1985, 223, 232; vgl. auch *Epstein/King*, U Chi L Rev 2002, 1, 19 Fn. 52 m.w.N.

[26] *Diekmann*, Sozialforschung 2012, 44: „Eine von vielen Fachleuten für wahr gehaltene plausible Erklärung oder Hypothese kann sich überraschend als falsch herausstellen.“

[27] *Lazarsfeld*, Publ Opin Q 1949, 377, 380: “Obviously something is wrong with the entire argument of ‘obviousness.’ It should really be turned on its head. Since every kind of human reaction is conceivable, it is of great importance to know which reactions actually occur most frequently and under what conditions”.

[28] *Fleischer*, ZGR 2008, 185, 188; ähnl. *Fraidin*, UC Davis L Rev 2004, 1, 67: “empirical psychology is superior to judicial intuition” (mit ausf. Begr.).

[29] UMAG v. 22.9.2005, BGBl. I S. 2802.

[30] Begr. Art. 1 Nr. 5 UMAG, BT-Drs. 15/5092 v. 14.3.2005, 15, 17, 20.

[31] Begr. Art. 1 Nr. 5 UMAG, BT-Drs. 15/5092 v. 14.3.2005, 13.

auf systematische Studien über die Häufigkeit der Hinterlegung berief sich der Gesetzgeber hier auf die Lebenserfahrung zahlreicher Marktteilnehmer, die zwar hohe Plausibilität, aber keine wissenschaftliche Bewährung beanspruchen konnte. Es könnte immerhin sein, dass einzelne Wirtschaftssektoren aus unbekannten Gründen an der Hinterlegung festgehalten, gleichwohl aber keine Möglichkeit (oder kein Interesse) gehabt hätten, sich im Gesetzgebungsverfahren zu äußern. Diese wohl fernliegende Gefahr hätte für sich genommen vielleicht nicht den Aufwand einer systematischen Erhebung gerechtfertigt, warnt aber vor allzu großem Vertrauen auf anekdotische Evidenz – sei sie noch so plausibel. Schließlich ist anekdotische Evidenz durch ihre subjektive Einfärbung immer auch ein Einfallstor (ob bewusst oder unbewusst genutzt) für Partikularinteressen.[32] Beispielsweise ging die Änderung des § 131 AktG durch Art. 1 Nr. 9 UMAG „auf Klagen aus der Praxis zurück, nach denen Auskunfts- und Rederecht häufig von einigen wenigen Aktionären missbraucht werden".[33]

Dass Lebenserfahrung unmittelbar an Erkenntniswert verliert, wenn bereits empirische Erkenntnisse vorliegen, wurde im UMAG-Gesetzgebungsverfahren ebenfalls deutlich, nämlich bei der Einführung des Freigabeverfahrens in § 246a AktG durch Art. 1 Nr. 23 UMAG. Hier berief sich der Gesetzgeber zwar scheinbar ebenso auf anekdotische Evidenz, nämlich auf das „Grundanliegen der Regierungskommission Corporate Governance und des 63. Deutschen Juristentages, missbräuchliche Ausübungen des Anfechtungsrechtes zu Lasten der Gesellschaft zu beschränken."[34] Tatsächlich aber hatte der zuständige Gutachter zum 63. Deutschen Juristentag eine empirische Studie „zur Beschlusskontrolle im Aktienrecht" durchgeführt und relevante Rechtstatsachen über die „Ausübungen des Anfechtungsrechts zu Lasten von Gesellschaften" systematisch erhoben.[35] Daher stützte sich der UMAG-Entwurf an jener Stelle implizit auf empirische Erkenntnis statt auf die anekdotische Evidenz der „Klagen aus der Praxis" – die doch hinsichtlich missbrauchter Anfechtungsrechte noch viel lauter gewesen waren als hinsichtlich missbrauchter Auskunftsrechte.

Strenger als die Rechtswissenschaft wahren die empirischen Disziplinen diesen Vorrang der kumulativen vor der anekdotischen Evidenz. Letztere kann systematische Studien allenfalls anregen[36] und hat insoweit – aber nur

---

[32] So auch *Rachlinski*, Cornell L Rev 2011, 901, 919: "Anecdotes come from somewhere, and not all anecdotes take hold. The anecdotes that support tort reform, in particular, seem to be promoted actively by the business community and the insurance industry."

[33] Begr. Art. 1 Nr. 5 UMAG, BT-Drs. 15/5092 v. 14.3.2005, 17.

[34] Begr. Art. 1 Nr. 5 UMAG, BT-Drs. 15/5092 v. 14.3.2005, 29.

[35] Dazu oben § 1 bei Fn. 291.

[36] *Garrn*, Rationalität 1986, 113 m.w.N. stellt in Fn. 158 fest, „dass jede sozialempirische Forschung ihrerseits auf einem Vorverständnis von sozialer Wirklichkeit beruht, das seine Ba-

insoweit – auch eine wissenschaftstheoretische Berechtigung (vgl. unten bei Fn. 162). Im Übrigen baut jeder empirische Erkenntnisprozess auf vorheriger empirischer Erkenntnis auf.

2. Kumulative Evidenz und empirische Erkenntnis

Um den empirischen Forschungsstand zu ermitteln, war traditionell eine Reihe von Quellen zu konsultieren,[37] die aber durch die fortschreitende Verbreitung digitaler Informationsquellen stark an Bedeutung verloren haben. Mittlerweile sind fast alle Zeitschrifteninhalte mindestens zurück bis zum Zweiten Weltkrieg digital indiziert, mit Kurzzusammenfassung (*abstract*) und oft sogar im Volltext abrufbar. Dasselbe gilt für viele Monographien, Lehr- und Handbücher, so dass eine Recherche in einschlägigen Datenbanken – allgemeinen[38] sowie fachspezifischen[39] – meist schon ein recht vollständiges Bild des Forschungsstands vermittelt.

Die empirische Literaturrecherche ähnelt daher weitgehend der juristischen. Eine Besonderheit allerdings zeichnet empirische Literaturdatenbanken aus: Sie indizieren nicht nur die eigentlichen Fachtexte, sondern auch deren Verweise untereinander. Dadurch wird es zum einen möglich, die Bedeutung und Wahrnehmung eines Textes an der Zahl anderer Texte zu messen, die ihn zitieren. Zum anderen ist auch eine vorwärtsgerichtete Verweissuche möglich: Ausgehend von einem Fachtext lassen sich nicht nur *frühere* Veröffentlichungen zum selben Thema auffinden, sondern mit der „Zitiert durch"-Funktion auch spätere.

Oft erfordert oder bedingt die Literaturrecherche eine Zusammenführung früherer Forschung, die je nach den Ambitionen des Forschers zu einer ganz eigenständigen Forschungsleistung auswachsen kann. Da solche Forschungs*synthese* denklogisch vorhandene Forschung voraussetzt, gehe ich erst am Ende dieses Abschnitts näher darauf ein (VII.). Was damit aber schon jetzt klar werden sollte, ist der spiralförmige Ablauf empirischer Forschung:[40] Die Recherche ist ebenso sehr der Anfang neuer wie die Fort-

---

sis [.] weitgehend im Bereich der Alltagserfahrung hat."; *Neumann/Krieger*, Clin L Rev 2003, 349, 380: "Anecdote does have some value, but primarily as illustration and not as proof. Used for illustration, it helps the data come alive, but that role is supplementary."; *Levine*, U Fla J L Pub Pol'y 2006, 283, 300 f.

[37] Vgl. *Bortz/Döring*, Forschungsmethoden 2006, 47 ff.

[38] Insbesondere SSRN (papers.ssrn.com), Google Scholar (scholar.google.com), JSTOR (www.jstor.com) und EBSCO (search.ebscohost.com).

[39] Beispielsweise APA PsycNet (psycnet.apa.org) für psychologische und AEA EconLit (www.aeaweb.org/econlit) sowie RePEc IDEAS (ideas.repec.org) für volkswirtschaftliche Veröffentlichungen.

[40] *Lercher* in: Barta u.a., Rechtstatsachenforschung 2009, 205, 215 nennt sie „einen iterativen Prozess".

setzung früherer Forschung, und die Synthese ist ebenso sehr ein Abschluss alter wie ein Ansatz zu neuer Forschung.[41] Synthese und Recherche gehen also ineinander über, und die Wahl eines ersten Schritts für eine sequentielle Darstellung ist hochgradig willkürlich.

II. Planung

Ist der Erkenntnisstand zu einer bestimmten (vorab formulierten oder während der Recherche entwickelten) Frage bekannt, sind die Schritte zu planen, um diese Frage zu beantworten. Diese Phase ist regelmäßig die wichtigste und aufwändigste im gesamten Forschungsablauf. An ihrem Ende steht ein Studienplan (*design*), der die Durchführung der Studie im Detail anleitet. Besonders fünf Aspekte sind dabei relevant: Worüber soll die Studie etwas aussagen? (dazu 1.) Was soll die Studie erreichen? (2.) Wie soll die Studie messen? (3.) Woher soll die Studie ihre Messdaten gewinnen? (4.) Wieviele Messdaten? (5.) Wie sollen die Messdaten ausgewertet werden? (6.)

1. Gültigkeitskriterium: Eindeutigkeit oder Übertragbarkeit

Zu Beginn der Studienplanung steht eine folgenreiche Entscheidung, die so wichtig ist und so oft missverstanden wird, dass ich ihr unter § 3 ein eigenes Kapitel widme. An dieser Stelle führe ich deshalb nur die wichtigsten Begriffe und die beiden Idealtypen ein.

Empirische Studien lassen sich grundsätzlich anhand zweier Arten von Gültigkeit beurteilen, nämlich anhand ihrer inneren und ihrer äußeren Gültigkeit, die auch interne und externe Validität genannt werden (von lat. *validus*, wirksam bzw. kräftig)[42]. Da beide Kriterien einander regelmäßig zuwiderlaufen, liegt es nahe, sie an zwei Enden einer Messlatte zu verorten. Das ist allerdings insoweit irreführend, als eine Verringerung der einen Gültigkeit nicht zwangsläufig die andere erhöht; vielmehr führen unterschiedliche empirische Methoden zu ganz verschiedenen Gewichtungen (dazu unten § 3 B.IV.1.). Daher erscheint es angemessener, innere und äußere Gültigkeit als zwei separate Kategorien zu verstehen, die aber miteinander wechselwirken.[43]

---

[41] *Chui* in: McConville/Chui, Methods 2007, 46, 52: "The end of one research cycle is the start of another".

[42] Dudenredaktion, Fremdwörterbuch 2000, 1386 („valid"); *Harper*, Etymology 2014, Stichwort "valid"; *Köbler*, Lateinisches Wb. 2009, Stichwort „validus"; im Folgenden vermeide ich den Begriff „Validität", weil er in der Testtheorie als Fachbegriff mit einer etwas anderen Bedeutung etabliert ist – dazu etwa *Sedlmeier/Renkewitz*, Forschungsmethoden 2008, 76 ff.; *Epstein/King*, U Chi L Rev 2002, 187 ff.

Jede der beiden Gültigkeiten misst gewissermaßen die Aussage*kraft* (vgl. *validus*) der Studie, gibt also an, worüber die Studie überhaupt Aussagen machen kann. Eine Studie mit hoher äußerer Gültigkeit kann unmittelbar etwas über den für die Forschungsfrage relevanten Lebenssachverhalt sagen, sie ist also leicht *übertragbar.*[44] Eine Studie mit hoher innerer Gültigkeit dagegen kann unmittelbar etwas über die für die Forschungsfrage relevanten Wirkungsbeziehungen sagen, sie kann also *eindeutig* Ursachen und Wirkungen unterscheiden.[45] Innere Gültigkeit ist damit synonym für die Eindeutigkeit (i.S. nur „einer Deutung“smöglichkeit der Ursachenzusammenhänge), äußere Gültigkeit synonym für die Übertragbarkeit der aus einer Studie gewonnenen Erkenntnis.[46]

Eindeutigkeit als Maß für die Unterscheidbarkeit von Ursache und Wirkung setzt (wie in der juristischen Kausalitätslehre) voraus, dass die vermeintliche Ursache als *conditio sine qua non* der beobachteten Wirkung erkannt werden kann. Theoretisch muss also dieselbe Situation in zwei Varianten verglichen werden – einmal ohne (*sine qua*) und einmal mit der vermeintlichen Ursache (*conditio*) –, um festzustellen, ob die Ursache eintritt oder nicht (*non*). In der Realität kann aber immer nur eine der beiden Situationen gleichzeitig eintreten, ihre kontrafaktische Alternative (*counterfactual*) lässt sich deshalb nie direkt beobachten. Daraus folgt,

> „dass kausale Wirkungen nicht direkt beobachtbar, sondern immer nur unter bestimmten Annahmen empirisch *erschlossen* werden können. Das fundamentale Problem jedweden Kausalschlusses besteht in der offensichtlichen Unmöglichkeit, dieselbe Person unter sonst gleichen Bedingungen *gleichzeitig* alternativen Rahmenbedingungen ausgesetzt zu sehen und damit ihre empirische [sic] Reaktionen auf unterschiedliche Handlungskontexte direkt erfassen zu können.“[47]

---

[43] Auch die geistigen Väter dieser Konzepte stellen beide Gültigkeiten grundsätzlich nebeneinander, betonen aber deren Wechselwirkungen: *Shadish/Cook/Campbell*, Causal Inference 2002, 37 ff.

[44] *Hussy/Schreier/Echterhoff*, Forschungsmethoden 2010, 131: „wenn die Ergebnisse der Untersuchung übertragbar sind“; zu eng dagegen (ausf. Begr. unten § 3 A.III.) *Bortz/Döring*, Forschungsmethoden 2006, 504: „wenn die Untersuchungsergebnisse auf andere, vergleichbare Personen, Orte oder Situationen generalisierbar sind.“

[45] *Hussy/Schreier/Echterhoff*, Forschungsmethoden 2010, 131: „wenn Veränderungen in der AV *ausschließlich* auf die Variation der UV zurückgeführt werden können.“; *Bortz/Döring*, Forschungsmethoden 2006, 504: „wenn die Untersuchungsergebnisse eindeutig für oder gegen die Hypothese sprechen und Alternativerklärungen unplausibel erscheinen.“

[46] Analysen mit hoher Übertragbarkeit bzw. Eindeutigkeit lassen sich auch mit den Begriffen „Totalanalyse“ (z.B. *Röhl*, Rechtssoziologie 1987, 70) bzw. „Partialanalyse“ (z.B. *Engel/Schön* in: Engel/Schön, Proprium 2007, IX, XII) kennzeichnen, doch suggerieren diese Begriffe ein Spezialitätsverhältnis, das nicht existiert.

[47] *Gangl/DiPrete*, KZfSS Sonderheft 2004, 396, 399; *Legewie*, KZfSS 2012, 123, 127: dass „per Definition nicht beide beobachtbar sind“, sei „das *fundamentale Problem der Kausalanalyse*“; *Ho/Rubin*, Ann Rev L Soc Sci 2011, 17, 21 fügt hinzu: “causal inference can

Allenfalls lassen sich sehr ähnliche Situationen beobachten, aber erstens sind in der komplexen Lebenswelt die Auswirkungen der verbleibenden Unterschiede (und seien sie noch so unscheinbar) kaum abschätzbar, und zweitens kann es immer übersehene oder unbeobachtbare Unterschiede zwischen den Situationen geben. Um derartige Unterschiede als Alternativerklärungen auszuschließen, muss der Forscher wohl oder übel in das Geschehen eingreifen und die Vergleichssituationen gezielt konstruieren.[48] Eindeutigkeit setzt deshalb gezielte *Experimente* unter ansonsten gleichen Bedingungen (lat. *ceteris paribus*) voraus. Wenn der eindeutige Ursachenrückschluss dagegen weniger wichtig und es stattdessen „sehr wichtig ist, die Ergebnisse auf das Alltagsleben zu generalisieren, dann sollte man versuchen, möglichst nicht in die Beobachtungssituation einzugreifen."[49] Übertragbarkeit verlangt deshalb nach möglichst berührungsfreier *Beobachtung*. Diese Differenzierung ist verwandt, aber nicht synonym, mit der Unterscheidung von reaktiven und nichtreaktiven Erhebungsmethoden:[50] Während letztere sich aus der Teilnehmerperspektive unterscheiden – danach, ob der Proband weiß, dass er an einer Studie teilnimmt – unterscheiden sich Beobachtung und Experiment aus der Studienleiterperspektive. Mithin gibt es reaktive und nichtreaktive Beobachtungsstudien ebenso wie reaktive und nichtreaktive Experimente.

Auf Einzelheiten und Zwischenschattierungen der beiden Idealtypen von Beobachtung und Experiment geht erst das nächste Kapitel näher ein. Vorliegend interessiert nur ihre praktische Relevanz für die Studienplanung: Soll eine Studie möglichst übertragbar sein, ist sie eher beobachtend anzulegen und muss vermeiden, das zu untersuchende Phänomen durch Eingriffe des Studienleiters unnötig zu stören, während solche Eingriffe geradezu Voraussetzung dafür sind, eindeutig Ursachen erkennen zu können.[51] Daraus erklärt sich, dass innere und äußere Gültigkeit einander regelmäßig widersprechen und dass zu Beginn jeder Studie beide Gültigkeiten gegeneinander abgewogen werden müssen.[52] Die Betonung liegt dabei

therefore be conceived of as a missing data problem."; ähnl. schon *Epstein/King*, U Chi L Rev 2002, 1, 37; *Harrison/List*, J Econ Lit 2004, 1009, 1014; *Epstein/Martin* in: Cane/Kritzer, Hdb Empirical 2010, 901, 903.

[48] *Holland*, J Am Stat Assoc 1986, 945, 959: "the motto NO CAUSATION WITHOUT MANIPULATION"; vgl. auch *Ho/Rubin*, Ann Rev L Soc Sci 2011, 17, 22.

[49] *Sedlmeier/Renkewitz*, Forschungsmethoden 2008, 108.

[50] Zu dieser Unterscheidung krit. *Diekmann*, Sozialforschung 2012, 195 f., 629.

[51] Zu beachten aber *Shadish/Cook/Campbell*, Causal Inference 2002, 34: "Validity is a property of inferences. It is *not* a property of designs or methods […] No method guarantees the validity of an inference."

[52] Vgl. *Walker/Willer* in: Webster/Sell, Experiments 2007, 51 ("internal validity and external validity are in opposition."); ausf. zu "Tradeoffs and Priorities" *Shadish/Cook/Campbell*, Causal Inference 2002, 96 ff.; *Lawless/Robbennolt/Ulen*, Methods 2010, 47; prägnant *Diek-*

auf „zu Beginn", weil die Abwägung zwischen Eindeutigkeit und Übertragbarkeit vor der Datenerhebung erfolgt sein muss. Keine noch so fortgeschrittene statistische Auswertung kann *beobachtete* Daten auf bestimmte Ursachen zurückführen,[53] und keine noch so geschickte Argumentation kann Erkenntnisse aus *experimentell* manipulierten Lebenssachverhalten auf nicht manipulierte eins-zu-eins übertragen.

2. Erkenntnisinteresse: beschreibend, schließend oder erkundend

Die zweite Weichenstellung im Rahmen der Studienplanung beeinflusst vor allem die Wahl der Datenquelle und die spätere Datenauswertung. Sie betrifft das Interesse, das hinter der systematischen Datenerhebung steht. Herkömmlich werden vier Erkenntnisinteressen unterschieden:[54]

Die wohl meisten Studien der empirischen Sozialforschung verfolgen ein *beschreibendes* Erkenntnisinteresse,[55] fassen also Daten über einen Lebenssachverhalt in einer Weise zusammen, die das Gesamtbild hervortreten lässt. Das ist beispielsweise das Gemeinsame der quantitativen Rechtstatsachenforschung, das sie zu einer abgrenzbaren empirischen Richtung vereint (§ 1 Fn. 258).

Ein großer Teil der empirischen Forschung verfolgt ein *schließendes* Erkenntnisinteresse und will aus einer Stichprobe bestimmte Regelmäßigkeiten in der Grundgesamtheit erkennen. Eine solche „Beschäftigung mit sozialwissenschaftlichen Gesetzmäßigkeiten, mit allgemeinen kausalen Zusammenhängen zwischen zwei sozialen Phänomenen" ist in Deutschland „sowohl in der juristischen Praxis als auch in der Rechtswissenschaft bisher nicht sonderlich stark ausgeprägt",[56] kennzeichnet aber sowohl die Ereignisstudienforschung (§ 1 G.II.) als auch die US-amerikanische empirische Rechtsforschung allgemein.[57]

Ein *evaluierendes* Erkenntnisinteresse verfolgt vor allem die empirische Anwendungsforschung (§ 1 bei Fn. 161) durch „die Ermittlung der Wirk-

---

mann, Sozialforschung 2012, 69: „Leider gilt auch im mühsamen Geschäft der empirischen Sozialforschung das Prinzip: ‚There is no such thing like a free lunch!'"

[53] *Röhl*, Rechtssoziologie 1987, 115; *Petersen/Goerg* in: Towfigh/Petersen, Methoden 2010, 201, 202; vgl. auch das Zitat bei *Ho/Rubin*, Ann Rev L Soc Sci 2011, 17, 20 m.w.N.: "Without … strong design, no amount of econometric or statistical modeling can make the move from correlation to causation persuasive".

[54] *Diekmann*, Sozialforschung 2012, 33 ff.; ohne das evaluierende Erkenntnisinteresse (vgl. unten bei Fn. 64) auch *Chui* in: McConville/Chui, Methods 2007, 46, 50 bei Fn. 23.

[55] *Diekmann*, Sozialforschung 2012, 36: „Wahrscheinlich verfolgt der Großteil aller Untersuchungen in der Sozialforschung primär deskriptive Ziele."

[56] *Petersen*, Staat 2010, 435.

[57] *Epstein/Martin* in: Cane/Kritzer, Hdb Empirical 2010, 901, 913: "For the vast majority of empirical legal projects, [.] making inferences [= schließende Erkenntnis] […] is the goal."

samkeit oder Unwirksamkeit praktisch-politischer oder sozialplanerischer Maßnahmen bezüglich eines oder mehrerer Erfolgskriterien.“[58] Solche Evaluationsforschung wurde schon an früherer Stelle dieser Arbeit als maßgebliche rechtspolitische Schnittstelle zur empirischen Forschung identifiziert (§ 1 F.II.).

Das vierte Erkenntnisinteresse ist das *erkundende*. Studien können zu einer Forschungsfrage neue Ideen und Perspektiven inspirieren, ohne zu ihrer Beantwortung unmittelbar beizutragen. Oft werden Studien mit erkundendem Erkenntnisinteresse allerdings als bloße Pilotstudien (Vorstudien) durchgeführt,[59] die „vorzugsweise qualitative Methoden“ nutzen,[60] um Hypothesen (dazu gleich 6.) für nachfolgende Studien mit schließendem Erkenntnisinteresse aufzustellen.

Anders als Lehrbücher bisweilen nahelegen,[61] begründen die verschiedenen Richtungen des Erkenntnisinteresses nicht prinzipiell unterschiedliche Arten von empirischer Forschung, sondern akzentuieren allenfalls die grundlegende Dichotomie zwischen Beobachtung und Experiment, die im vorangegangenen Abschnitt eingeführt wurde. Welches Erkenntnisinteresse verfolgt wird, hängt maßgeblich vom Stand der Erkenntnis ab: Je weniger über ein Thema bekannt ist, desto mehr Gewicht wird auf beschreibende oder bloß erkundende Studien zu legen sein,[62] während umgekehrt der „nächste logische Schritt“ nach solchen Studien das schließende Erkenntnisinteresse bedient.[63]

Die unterschiedlichen Erkenntnisinteressen bedürfen jeweils unterschiedlicher Methoden der Datenauswertung und werden deshalb dort (IV. vor 1.) näher besprochen. Dabei geht das evaluierende Erkenntnisinteresse im schließenden auf, da es methodisch nur einen „Sonderfall hypothesenprüfender [d.h. schließender] Studien“ mit „direktem Anwendungsbezug auf versuchsweise durchgeführte oder bereits realisierte Maßnahmen“ darstellt.[64]

3. Operationalisierung: Variablen und ihr Skalenniveau

Im dritten Planungsschritt müssen die Begriffe und Konzepte, die der Forschungsfrage unterliegen, in empirisch beobachtbare Messgrößen (Indika-

---

[58] *Diekmann*, Sozialforschung 2012, 37; ausf. *Bortz/Döring*, Forschungsmethoden 2006, 95 ff. (Kap. 3 „Besonderheiten der Evaluationsforschung“); *Arndt*, Empirie 2008, 80 ff.

[59] *Chui* in: McConville/Chui, Methods 2007, 46, 50.

[60] *Diekmann*, Sozialforschung 2012, 33 f. (Zitat auf 34).

[61] Z.B. *Leary*, Methods 2008, 22, der empirische Forschung in die Kategorien *descriptive*, *correlational*, *experimental* und *quasi-experimental* aufteilt

[62] *Bortz/Döring*, Forschungsmethoden 2006, 50; *Diekmann*, Sozialforschung 2012, 33 f.

[63] *Chui* in: McConville/Chui, Methods 2007, 46, 51.

[64] *Diekmann*, Sozialforschung 2012, 38.

toren) übersetzt (operationalisiert) werden.[65] Dazu werden verschiedene Arten von Variablen definiert: üblicherweise eine oder mehrere *unabhängige Variablen*, die vermutete Ursachen quantifizieren, sowie eine *abhängige Variable*, die die vermutete Folge quantifiziert.[66] Die Ausprägungen der unabhängigen Variablen werden gezielt gesucht oder hergestellt und sind damit *exogen*, während die Ausprägungen der abhängigen Variable sich *endogen* daraus ergeben.[67] Weil die unabhängigen Variablen folglich zur Erklärung der abhängigen beitragen, heißen sie auch erklärende Variablen oder Prädiktoren.[68] Daneben sind eine Vielzahl von *Moderatorvariablen* denkbar für Größen, die die Ursachenbeziehung mitbeeinflussen; werden sie ebenfalls erhoben, spricht man von *Kontrollvariablen*, andernfalls von *Störvariablen*.[69]

Variablen lassen sich auf vier verschiedenen *Skalenniveaus* messen, die den Ausschlag darüber geben, welche Arten der Datenauswertung möglich sind.[70] Daten auf einem höheren Skalenniveau enthalten mehr Informationen, können aber jederzeit auf ein niedrigeres Skalenniveau überführt werden, indem ein Teil ihres Informationsgehalts vernichtet wird.

Das niedrigste Skalenniveau hat die Nominalskala, von lat. *nomen* (Name). Aus der Bezeichnung wird bereits deutlich, dass eine Nominalskala lediglich „Namen“ bzw. Bezeichnungen enthält. Die einzige Beziehung, in die sich solche Daten setzen lassen, ist die der Gleichheit oder Ungleichheit, weitere Informationen vermitteln sie dagegen nicht. Als Beispiel könnte etwa eine Rechtsformenskala dienen, auf der sich jede Gesellschaft eindeutig einordnen lässt, die aber nicht mehr und nicht weniger ergibt, als dass zwei Gesellschaften die gleiche oder unterschiedliche Rechtsformen haben.

---

[65] *Beller*, Forschen lernen 2008, 17, 29; *Bortz/Döring*, Forschungsmethoden 2006, 63: „Eine operationale Definition standardisiert einen Begriff durch die Angabe der Operationen, die zur Erfassung des durch den Begriff bezeichneten Sachverhaltes notwendig sind, oder durch Angabe von messbaren Ereignissen, die das Vorliegen dieses Sachverhaltes anzeigen (Indikatoren).“; juristische Beispiele bei *Lercher* in: Barta u.a., Rechtstatsachenforschung 2009, 205, 208.

[66] *Abelson*, Principled Argument 1995, 190: “One might say that general cause and general effect are each *instantiated* by particular cause and particular effect.”

[67] Wortherkunft von griech. ἔξω *exo*, außen, bzw. ἔνδον *endon*, innen (seinerseits von *en* und *domo*, im Haus) sowie γένος *genos*, Geburt (Dudenredaktion, Fremdwörterbuch 2000, 391 bzw. 427; *Harper*, Etymology 2014, Stichworte “endogenous”, “exogenous”, “endo-”, “exo-”, “-gen”).

[68] Von lat. *prae* und *dicere*, vorher sagen (*Dudenredaktion*, Fremdwörterbuch 2000, 1074; *Harper*, Etymology 2014, Stichwort “predict”).

[69] *Beller*, Forschen lernen 2008, 16 f.; *Bortz/Döring*, Forschungsmethoden 2006, 3.

[70] Statt vieler *Diekmann*, Sozialforschung 2012, 285 ff. (Übersichtstabelle 291), 687.

Das nächsthöhere Skalenniveau hat die Ordinalskala, von lat. *ordo* (Reihenfolge). Mithin enthält die Ordinalskala nicht nur Informationen über die Identität, sondern auch über die Reihen- bzw. Rangfolge. Ordinalskalierte Daten lassen sich nicht nur ins Verhältnis der Gleichheit oder Ungleichheit setzen, sondern auch ins Größer-als- oder Kleiner-als-Verhältnis. Als Beispiel könnte etwa eine Gläubigerschutzskala dienen, auf der sich jede Rechtsform zwischen der GbR und der AG einordnen und gegenüber jeder anderen als gläubigerfreundlicher oder weniger gläubigerfreundlich charakterisieren lässt.

Das zweithöchste Skalenniveau hat die Intervallskala, von lat. *intervallum* (Zwischenraum). Zusätzlich zu den bisherigen Informationen lässt sich aus ihr auch noch ablesen, welcher Zwischenraum bzw. Abstand zwischen den nach Rängen geordneten Werten besteht. Daher lassen sich solche Daten auch in Additionen und Subtraktionen einsetzen. Als Beispiel könnte etwa ein Zeitstrahl dienen, auf dem die gesetzliche Kodifizierung jeder Rechtsform mit einer Jahreszahl versehen ist, so dass sich ablesen lässt, *wieviel* früher eine Rechtsform im Vergleich zu einer anderen kodifiziert wurde.

Das höchste Skalenniveau hat die Proportionalskala, von lat. *proportio* (Verhältnis). Diese Skala hat im Gegensatz zu allen anderen Skalen einen absoluten Nullpunkt und erlaubt deshalb auch Aussagen über Verhältnisse. Die auf ihr befindlichen Daten lassen sich also unmittelbar in Multiplikationen und Divisionen verwenden. Als Beispiel könnte etwa eine Stammkapitalskala dienen, auf der sich das Stammkapital einer GmbH ablesen und zugleich als Vielfaches oder Teil des Stammkapitals jeder anderen GmbH kennzeichnen lässt.

4. Datenquelle und Erhebungsart: Längs- oder Querschnitt

Ist definiert, welche Variablen auf welchem Skalenniveau zu erheben sind, bleibt zu überlegen, welche Datenquelle für die Studie herangezogen wird.[71] Hier haben Computer und Internet zu einem enormen Zuwachs der Möglichkeiten geführt,[72] der aber auch stärker als bisher zu Überlegungen darüber zwingt, welche Daten überhaupt für die Forschungsfrage geeignet sind. Dabei wirken sich sowohl das gewählte Gültigkeitskriterium als auch das Erkenntnisinteresse erneut aus:

Studien mit hoher innerer Gültigkeit setzen Daten aus einem Lebensbereich voraus, in den der Forscher manipulierend eingreifen kann. Legt er

---

[71] Vgl. z.B. *Epstein/King*, U Chi L Rev 2002, 1, 99 ff. (Abschnitt VIII).

[72] *Diekmann*, KZfSS Sonderheft 2004, 8 ff.; *Rachlinski*, Cornell L Rev 2011, 901, 909: “The combination of the Internet and the desktop computer has liberated the modern empirical legal researcher.”; vgl. noch unten Fn. 299.

dagegen mehr Wert auf die äußere Gültigkeit, muss er seine Daten aus einem Lebensbereich schöpfen, den er ungehindert beobachten kann; rechtliche Verbote ziehen hier oft eine Grenze.[73] (Nutzt er stattdessen bestehende Daten, muss er die Zuverlässigkeit der Datenquelle abschätzen können.)

Studien mit erkundendem Erkenntnisinteresse sind in der Wahl ihrer Datenquelle unbeschränkt. Verfolgt die Studie dagegen ein beschreibendes Erkenntnisinteresse, muss der relevante Lebensbereich vollständig beobachtet werden, während für ein schließendes Erkenntnisinteresse die Ziehung einer Stichprobe genügt.[74]

In jedem Fall ist weiterhin zu überlegen, ob die Studie einen Quer- oder Längsschnitt erheben oder beides kombinieren soll.[75] In Beobachtungsstudien bedeutet Längsschnitt (*longitudinal design*) die wiederholte Beobachtung desselben Beobachtungsobjekts, die allen Zeitreihenerhebungen (Trenddesigns) zugrunde liegt, Querschnitt dagegen (*cross-sectional design*) eine Momentaufnahme verschiedener Beobachtungsobjekte. In Experimentalstudien bedeutet Längsschnitt (*within subject design*), dass jeder Teilnehmer mehrere Versuchsbedingungen nacheinander absolviert, Querschnitt dagegen (*between subjects design*), dass jeder Teilnehmer nur einer von mehreren Versuchsbedingungen zugeteilt wird.[76] In Längsschnittstudien lassen sich Eigenarten der Beobachtungsobjekte durch Mittelwertsbildung herausrechnen, weshalb sie auch bei kleinerer Stichprobe weniger störende Datenstreuung aufweisen. Zugleich haben sie den Nachteil, dass zeitlich nachfolgende Beobachtungen immer von den früheren abhängen,[77] was aufwändige Vorkehrungen bei Studienplanung und Datenauswertung erfordert. Längs- und Querschnitt lassen sich kombinieren (Zeitreihe mehrerer Beobachtungsobjekte), dann spricht man von einer Panelstudie.[78]

---

[73] *Upmeier*, Fakten 2010, 148; *Röhl*, Rechtssoziologie 1987, 106 mit der Kritik, dass „es leider noch nicht selbstverständlich ist, auch ein wissenschaftliches Interesse als rechtliches i.S. von § 299 ZPO zu behandeln.“; *Baldwin/Davis* in: Cane/Tushnet, Hdb Legal Studies 2003, 880, 893: “Legal researchers may find that it is not possible for them to examine certain subjects”.

[74] Zu Zufallsziehung und Repräsentativität der Stichprobe gleich IV. nach Fn. 116.

[75] Vgl. *Lawless/Robbennolt/Ulen*, Methods 2010, 34.

[76] Die englischen Bezeichnungen sind „auch im Deutschen“ üblich, so *Sedlmeier/Renkewitz*, Forschungsmethoden 2008, 150 f.

[77] *Lawless/Robbennolt/Ulen*, Methods 2010, 35: “examining effects over time is often messy”; zum möglichen Aufforderungseffekt in *within-subject*-Studien *Abelson*, Principled Argument 1995, 195.

[78] *Bortz/Döring*, Forschungsmethoden 2006, 447: „Stichprobe, die wiederholt […] befragt wird“; von lat. *pannus*, Stoff – übertragen auf die Pergamentliste der Geschworenen, später engl. *panel* für Spruchkörper allgemein (*Harper*, Etymology 2014, Stichwort “panel”).

5. Stichprobengröße: Effektgröße, Spezifität und Teststärke (power)

Soweit der Forscher seine Daten selbst erhebt – ob beobachtend oder experimentell – stellt sich die Frage, *wieviele* Daten. In der Abwägung zwischen Kosten und Nutzen entscheidet regelmäßig das Prinzip „So viele wie nötig, so wenige wie möglich". *Wie* viele nötig sind, ist Gegenstand diverser Faustregeln,[79] die aber nicht darüber hinwegtäuschen dürfen, dass die einzig richtige Antwort lautet: Es kommt darauf an. Nämlich auf Effektgröße, Spezifität und Teststärke:

Effektgröße (näher unten nach Fn. 115) ist die Größe des Unterschieds (relativ zur Datenstreuung), den man zu finden erwartet. Je kleiner diese erwartete Effektgröße ist, desto größere Stichproben werden benötigt, um den Effekt nachzuweisen.

Spezifität ist die Wahrscheinlichkeit, keinen Effekt zu beobachten, wo keiner existiert. Sie gibt also an, wie oft bei sehr häufiger Ziehung von Stichproben ein tatsächlich nicht existierender Unterschied auch in der Stichprobe ausbleibt. In allen anderen Fällen liegt ein *falsch-positives* Ergebnis vor, das als Fehler 1. Art oder α-Fehler bezeichnet wird. Per Konvention soll die langfristige α-Fehlerwahrscheinlichkeit (Signifikanzniveau) höchstens 5 % betragen (vgl. unten bei Fn. 157).

Die Teststärke (*power*), auch Sensitivität genannt, ist die Wahrscheinlichkeit, einen Effekt zu beobachten, wo er existiert. Sie gibt also an, wie oft bei sehr häufiger Ziehung von Stichproben ein tatsächlich existierender Unterschied auch in der Stichprobe auftaucht. In allen anderen Fällen liegt ein *falsch-negatives* Ergebnis vor, das auch als Fehler 2. Art oder β-Fehler bezeichnet wird. Per Konvention wird eine langfristige Teststärke von mindestens 80 % als ausreichend angesehen.

Spezifität und Teststärke beeinträchtigen sich gegenseitig, was anschaulich mit dem Bild eines Gerichtsverfahrens beschrieben wird:[80] Die einzige Richterin, die niemals einen Angeklagten zu Unrecht verurteilen kann, ist diejenige, die aus Prinzip *jeden* freispricht, während die einzige Richterin, die niemals einen Angeklagten zu Unrecht freisprechen kann, diejenige ist, die *jeden* verurteilt. Alle Richterinnen dazwischen können immer Fehler

---

[79] *Croson*, U Ill L Rev 2002, 921, 939 ("twenty to thirty independent observations for each treatment"); *VanVoorhis/Morgan*, Psi Chi J 2001, 139; *van Belle*, Rules of Thumb 2008, 27 ff. (Kap. 2); *Brewer*, Fl J Edu Res 1988, 5 ("rules-of-thumb should be invoked only if absolutely necessary"); speziell für Regressionen *Green*, Multivar Behav Res 1991, 499.

[80] Bspw. *Abelson*, Principled Argument 1995, 16; *Lawless/Robbennolt/Ulen*, Methods 2010, 228 f.; *Upmeier*, Fakten 2010, 147; *Diekmann*, Sozialforschung 2012, 714; umgekehrt (statistische Tests als Metapher für das Gerichtsverfahren) etwa *Rogers*, Mistakes 2001; *Häusermann*, ZStrR 2011, 194.

machen und müssen sich entscheiden, ob sie mehr zur Milde (Spezifität) oder zur Strenge (Teststärke) neigen wollen.[81]

Wenn Spezifität und Teststärke gegeneinander abgewogen sind und eine Erwartung hinsichtlich der Effektgröße beziffert werden kann, lässt sich errechnen, wie groß die Stichprobe sein muss.[82] Leider wird eine solche Teststärkenberechnung (*power analysis*) sehr selten vorab durchgeführt, weshalb sich in vielen Disziplinen herausgestellt hat, dass die Forschung jahrzehntelang mit so kleinen Stichproben gearbeitet hat, dass die gesuchten Effekte mit höherer Wahrscheinlichkeit verpasst als erkannt werden mussten.[83] Nun möchte man meinen, das sei ein Problem der Forscher, nicht aber der Rezipienten, denn ist ein Effekt erst einmal belegt, interessiert die „gigantische Verschwendung von forscherischem Enthusiasmus und materiellen Ressourcen für sinnlose Zufallsergebnisse"[84] nicht mehr. Allerdings haben Forscher bei der Veröffentlichung ihrer Ergebnisse (dazu unter VI.) einen erheblichen Anreiz, Nullergebnisse unter den Tisch fallen zu lassen und Resultate stattdessen schönzurechnen. Solche schöngerechneten Ergebnisse werden daher umso häufiger auftreten, je öfter Forscher Nullergebnisse finden – und damit beeinflusst die Wahl der Stichprobengröße durchaus die Qualität der veröffentlichten Forschungsergebnisse:

Aus Studien, deren Stichproben zu klein sind, um einen tatsächlich existierenden Effekt zu finden, können Studien werden, die einen nicht existierenden Effekt berichten, nur um überhaupt etwas berichten zu können.

6. Hypothesen und Auswertungsmethoden

Falls der Forscher – wie meist – aus einer Stichprobe verallgemeinern möchte, muss er schließlich konkrete Hypothesen über den Zusammenhang der definierten Variablen aufstellen. Das ist erforderlich, weil die schließende Statistik, die zur Verallgemeinerung aus Stichproben benötigt wird, auf dem Vergleich naiver (uninformierter) Wahrscheinlichkeitsverteilungen mit tatsächlich beobachteten Daten beruht. Daher liefert sie sinnlose und uninterpretierbare Ergebnisse, wenn die Wahl der Wahrscheinlichkeitsver-

---

[81] Zu einer ähnlichen Abwägung (Freiheit und Sicherheit im Grundrechtsschutz) vgl. *Hassemer*, Vorgänge 2002, 3/10; weitere anschauliche Metapher (Brandmelder) bei *Dubben/Beck-Bornholdt*, KZfSS Sonderheft 2004, 61, 66.

[82] Es gibt ganz unterschiedliche Effektarten, also auch nicht nur eine einheitliche Formel. Deswegen werden entsprechende Berechnungen praktisch nie ohne Software durchgeführt.

[83] Laienverständlich und ausf. *Ellis*, Effect Sizes 2010, xiv f. sowie 47 ff. (Kap. 3 und 4) m.w.N. insb. auf S. 76; ferner *Bortz/Döring*, Forschungsmethoden 2006, 602 ff.; *Beller*, Forschen lernen 2008, 110 ff.; Abhilfe bei *Shadish/Cook/Campbell*, Causal Inference 2002, 46 f.

[84] *Dubben/Beck-Bornholdt*, KZfSS Sonderheft 2004, 61, 73.

teilung in irgendeiner Weise von den beobachteten Daten beeinflusst wird.[85] Dem beugen vorab definierte Hypothesen vor.

Hypothesen sind Aussagen, die vier Voraussetzungen erfüllen: Erstens müssen sie sich „auf reale Sachverhalte, die empirisch untersuchbar sind" beziehen; zweitens müssen sie „allgemein gültige, über den Einzelfall oder ein singuläres Ereignis hinausgehende" (nomologische) All-Behauptungen aufstellen; drittens „muss zumindest implizit die Formalstruktur eines sinnvollen Konditionalsatzes (»Wenn-dann-Satz« bzw. »Je-desto-Satz«) zugrunde liegen"; und viertens muss dieser Konditionalsatz auch „potenziell falsifizierbar sein, d. h., es müssen Ereignisse denkbar sein, die dem Konditionalsatz widersprechen."[86]

Praktisch bedeutet das, dass Hypothesen so genau wie möglich (näher unten nach Fn. 283) eine Aussage darüber machen, wie sich die abhängige Variable ändert, wenn sich die unabhängigen Variablen ändern.[87] Diese Aussage stützt sich regelmäßig auf den Stand der bis dahin entwickelten Theorie (dazu unten VII.3.).

Zugleich mit der Hypothese sollten die Auswertungsmethoden festgelegt werden; in geringem Umfang kann die Wahl der Auswertungsmethode zwar noch von Eigenheiten der Daten beeinflusst werden,[88] aber die wesentlichen Kriterien für die Wahl einer Auswertungsmethode (Skalenniveau, Stichprobengröße, Abhängigkeiten, etc.) sind bereits vor der Datenerhebung bekannt.

III. Erhebung

Nach Abschluss der Studienplanung sind die Daten zu erheben. Dabei bedarf es einer durchdachten Systematik zur Aufzeichnung (Kodierung, von engl. *coding*) der Daten,[89] um sie später statistisch auswerten zu können.

---

[85] Denn beobachtete Daten haben eine „Wahrscheinlichkeit" von 100 %, vgl. *Duckworth*, Teach Stat 2006, 84, 86.

[86] *Bortz/Döring*, Forschungsmethoden 2006, 4; noch genauer (acht Voraussetzungen) *Atteslander*, Methoden 2010, 37.

[87] Im Detail lassen sich spezifische und unspezifische, gerichtete und ungerichtete, Zusammenhangs- und Unterschiedshypothesen unterscheiden (*Beller*, Forschen lernen 2008, 100); noch detaillierter *Bortz/Döring*, Forschungsmethoden 2006, 502 ff.

[88] Etwa, dass die Daten zensiert (gestutzt) sind; dafür schlug *Tobin*, Econometrica 1958, 24 eine mittlerweile etablierte Erweiterung der Regressionsrechnung vor (Tobit-Regression). Idealerweise sind die Vorstellungen über den Vorgang der Datenerzeugung (*data generating process*) natürlich so konkret, dass eine etwaige Zensierung schon bei der Wahl der Auswertungsmethode, gleichzeitig mit der Hypothesenbildung, berücksichtigt werden kann.

[89] Dazu Epstein/Martin in: Cane/Kritzer, Hdb Empirical 2010, 901, 911 f.; ausf. *Lawless/Robbennolt/Ulen*, Methods 2010, 165 ff. (Kap. 7).

Der genaue Ablauf der Datenerhebung hängt natürlich von den gewählten Methoden ab,[90] allerdings lässt sich allgemein feststellen:

> „Ist eine Untersuchung sorgfältig und detailliert geplant, dürfte ihre Durchführung keine besonderen Schwierigkeiten bereiten. Was aber durch Planung als potenzielle Störquelle nicht völlig ausgeschlossen werden kann, sind Fehler im eigenen Verhalten [des Forschers…] Die Literatur spricht in diesem Zusammenhang von Versuchsleiterartefakten."[91]

Mit Versuchsleitereffekten sind je nach Erhebungsmethode (beobachtend oder experimentell) Verhaltensweisen gemeint, die entweder den beobachteten Vorgang ungewollt beeinflussen,[92] oder den gewollt beeinflussten Vorgang nicht stets in gleicher Weise beeinflussen. So können bereits subtile körpersprachliche Signale gegenüber den Studienteilnehmern eine Erhebung drastisch verfälschen, was auch als „Kluger-Hans-Effekt" bezeichnet wird, nach dem berühmten Fall eines Pferdes, das angeblich rechnen konnte, aber tatsächlich nur auf kaum wahrnehmbare Signale in der Körpersprache der umstehenden Beobachter reagierte.[93]

Eine wichtige Art von Versuchsleiterartefakten sind die Aufforderungseffekte (*demand effects*), die dann entstehen können, wenn Studienteilnehmer den Zweck einer Untersuchung durchschauen und sich deshalb entweder besonders zweckkonform oder besonders widerspenstig verhalten, je nach ihrer Einstellung zum Versuch, seinem Zweck und der Person des Versuchsleiters. Durch diese Reaktivität entsteht eine weitere Störvariable, die eindeutige Ursachenrückschlüsse vereiteln kann.

Um solche Aufforderungseffekte zu vermeiden, steht nicht nur der aus der Medizin bekannte Doppelblindversuch zur Verfügung, sondern auch eine Reihe weniger aufwändiger Vorkehrungen,[94] die meist darauf hinauslaufen, das Ziel einer Studie gegenüber ihren Teilnehmern zu verschleiern (oder sie über ihre Teilnahme an einer Studie ganz im Unklaren zu lassen). Daher stellen sich auch forschungsethische und rechtliche Fragen, etwa nach Freiwilligkeit und Informiertheit als Voraussetzungen einer rechtfertigenden Einwilligung in Eingriffe in das Selbstbestimmungsrecht.[95]

---

[90] Für Experimente z.B. *Croson*, U Ill L Rev 2002, 921, 924 f.

[91] *Bortz/Döring*, Forschungsmethoden 2006, 81 f.; vgl. auch *Abelson*, Principled Argument 1995, 193 ff. ("procedural bias").

[92] Zu diesem sog. Hawthorne-Effekt unten § 3 bei Fn. 241.

[93] *Bortz/Döring*, Forschungsmethoden 2006, 82 m.w.N.; *Diekmann*, Sozialforschung 2012, 623 f. m.w.N. zieht die Verbindung zum sog. Pygmalion-Effekt selbsterfüllender Prophezeiungen; *Abelson*, Principled Argument 1995, 194 f. ("demand characteristics"); ähnl. Bsp. bei *Stanovich*, Psychology 2012, 85 ff. (Kap. 6).

[94] Näher *Bortz/Döring*, Forschungsmethoden 2006, 83 f.; *Hussy/Schreier/Echterhoff*, Forschungsmethoden 2010, 54 ff.

[95] Dazu *Eberbach/Schuler*, JZ 1982, 356; *Deutsch* in: Ehmann u.a., FG Weitnauer 1980, 297; *Bortz/Döring*, Forschungsmethoden 2006, 41 ff. (in Kap. 2.2); *Hussy/Schreier/Echter-*

IV. Auswertung

Zur Auswertung quantitativer Daten bedient sich die empirische Sozialforschung einer mathematischen Teildisziplin, die zusammen mit der Wahrscheinlichkeitsrechnung der Stochastik zugeordnet wird: der Statistik.

Nach ihrer Wortherkunft (von lat. *statisticum,* den Staat betreffend)[96] sollte die Statistik wohl eher Staatistik heißen, denn ihre Anfänge liegen in der Staatslehre des 18. Jahrhunderts, die durch den absolutistischen Verwaltungsapparat aufgrund zunehmender Datenmengen immer stärker quantifiziert wurde.[97] Nach Gründung der Königlichen Statistischen Gesellschaft in England 1835 wurde „Statistik" dann auf jegliche Datenmengen bezogen und bezeichnet heute ganz allgemein „alle Methoden der Analyse von Daten mit dem Ziel der Informationsbündelung".[98] Diese Auswertungsmethoden greifen heute durchweg auf statistische Software zurück.[99]

Die Statistik verfügt über zwei Teilgebiete mit unterschiedlichem, meist komplementärem, Zweck:[100]

Die *beschreibende Statistik* übersetzt große Datenmengen in handlichere Kennzahlen bzw. Parameter (von griech. παρά *para* und μέτρον *metron*, daneben messen).[101] Sie vollzieht also einen Abstraktionsschritt, indem sie Detailinformationen zugunsten des Überblicks über das Ganze aufgibt. Damit erfüllt sie zwei Funktionen: Soweit ein beschreibendes Erkenntnisinteresse verfolgt, also die relevante Grundgesamtheit vollständig beobachtet

---

hoff, Forschungsmethoden 2010, 41 ff. (Kap. 1.7); *Chui* in: McConville/Chui, Methods 2007, 46, 62 f.

[96] *Harper*, Etymology 2014, Stichwort "statistics"; Dudenredaktion, Fremdwörterbuch 2000, 1269: „wohl zu nlat. statisticus, staatswissenschaftlich"; *Köbler*, Etymologisches Wb. 1995, 383: „nlat. statistica, F., Staatslehre".

[97] Ausf. *Yule*, Statistics 1911, 2 m.w.N.: Statistik, "as the term is used by German writers of the eighteenth century […] meant simply the exposition of the noteworthy characteristics of a state […] After the commencement of the nineteenth century, however, the growth of official data was continuous, and numerical statements, accordingly, began more and more to displace the verbal descriptions of earlier days."; ähnl. *Diekmann*, Sozialforschung 2012, 94 f.; *Abelson*, Principled Argument 1995, 2 Fn. 1; *Krämer*, KZfSS Sonderheft 2004, 51, 53 hält die Homonymie für „unglücklichen Zufall".

[98] *Quatember*, Statistik ohne Formeln 2011, 10; *Yule*, Statistics 1911, 7.

[99] Überblick bei *Bortz/Döring*, Forschungsmethoden 2006, 751 f. (Anh. D), aber ohne einige gängige Programme, wie das kommerzielle Paket Stata (www.stata.com/links/resources-for-learning-stata), die anspruchsvolle aber freie Programmiersprache R (www.r-project.org) und die statistischen Funktionen von Microsoft Excel (vgl. *Hagen*, Statistik für Juristen 2005; *Matthäus*, Tests mit Excel 2007; *Quatember*, Statistik ohne Formeln 2011).

[100] Plastisch unterschieden im Titel des Lehrbuchs von *Weigand*, Statistik 2009; für juristische Fragen *Epstein/King*, U Chi L Rev 2002, 1, 24 ff., 34 ff. m. Bsp.

[101] *Harper*, Etymology 2014, Stichwort "parameter"; *Dudenredaktion*, Fremdwörterbuch 2000, 984 („para-"), 867 („-meter"); *Salsburg*, Statistics 2001, 16: "almost measurements".

wurde (Vollerhebung),[102] ist die beschreibende Statistik eine eigenständige und abschließende Form der Datenauswertung. Soweit dagegen ein schließendes Erkenntnisinteresse verfolgt und nur ein Teil der relevanten Grundgesamtheit beobachtet wurde (Stichprobenerhebung),[103] bereitet die beschreibende Statistik den Einsatz des zweiten statistischen Teilgebiets vor.

Das ist die *schließende Statistik*,[104] die wahrscheinlichkeitstheoretisch fundierte Schlüsse von den Parametern einer Stichprobe auf die Parameter der Grundgesamtheit ermöglicht. Diese Parameterschätzung ist vor allem deshalb nützlich, weil sie zur Überprüfung von Hypothesen eingesetzt werden kann, indem eine „Grundgesamtheit" fingiert wird. Um etwa zu beantworten, ob nominell unterkapitalisierte Gesellschaften höhere Risiken eingehen als voll eigenkapitalfinanzierte Gesellschaften, würde die schließende Statistik entsprechende Stichproben darauf überprüfen, ob ihre Parameter zur selben hypothetischen Grundgesamtheit passen, in der nominelle Kapitalisierung und Risikofreudigkeit proportional zueinander sind, oder ob sie zwei verschiedenen hypothetischen Grundgesamtheiten entstammen. Deshalb sind die beiden oft unterschiedlich akzentuierten oder gar voneinander abgegrenzten Funktionen der schließenden Statistik – Verallgemeinerung aus Stichproben und Testen von Hypothesen – letztlich äquivalent.

Als drittes Teilgebiet der Statistik wird oft die *erkundende Statistik* genannt, allerdings ist diese Dreiteilung irreführend, denn die erkundende Statistik unterscheidet sich nur im Erkenntnisinteresse, nicht dagegen in den Methoden, von den beiden Teilgebieten der Statistik. Sie ist eher eine besondere statistische Anwendung und daher nur kurz darzustellen.

1. Beschreibende (deskriptive) Statistik

Beschreibende Statistik macht unhandliche Datenmengen handhabbar. Eine Datenmenge besteht aus den bei zahlreichen Untersuchungsobjekten beobachteten Ausprägungen eines oder mehrerer Merkmale. Zu jedem Merkmal gibt es also eine empirische *Verteilung* seiner Ausprägungen, die die beschreibende Statistik in Parameter übersetzen kann.[105] Für ein einzelnes Merkmal sind die sog. Lage- und Streuungsparameter am wichtigsten, das

---

[102] Die deskriptive Statistik belegt also *singuläre* Sätze über die beobachtete Datenmenge (vgl. *Diekmann*, Sozialforschung 2012, 150), nicht wissenschaftliche Hypothesen, also „All-Sätze" (oben bei Fn. 86 sowie *Herberger/Simon*, Wissenschaftstheorie 1980, 344).

[103] Die Suche nach *allgemeinen* Gesetzmäßigkeiten ist immer eine Stichprobenerhebung: "an observed population is a 'sample' from possible histories" (*Epstein/Martin* in: Cane/Kritzer, Hdb Empirical 2010, 901, 910 Fn. 9).

[104] Synonyme: mathematische, analytische, beurteilende, induktive, inferentielle oder Inferenzstatistik, letztere von lat. *in* und *ferre*, herein tragen (*Harper*, Etymology 2014, Stichwort "infer"; *Köbler*, Lateinisches Wb. 2009; Dudenredaktion, Fremdwörterbuch 2000, 611).

[105] Anschaulich *Petersen/Goerg* in: Towfigh/Petersen, Methoden 2010, 201, 212 ff.

Verhältnis mehrerer Merkmale zueinander beschreiben sog. Zusammenhangsparameter.

Der wichtigste *Lageparameter* ist das Maß der zentralen Tendenz: Wo ist die Mitte der Daten? Je nach Skalenniveau (oben nach Fn. 70) sind verschiedene Maße der zentralen Tendenz denkbar: Nominalskalierte Daten erlauben definitionsgemäß nur eine Aussage über den häufigsten Wert (Modus). Ordinalskalierte Daten können zusätzlich durch den mittleren Wert (Median) charakterisiert werden, also denjenigen Wert, der gleich viele Elemente über und unter sich hat, wenn die Zahlenfolge nach Größe geordnet wird. Erst aus intervallskalierten Daten lässt sich außerdem noch ein arithmetisches Mittel (Mittelwert) errechnen, indem die Summe aller Elemente durch ihre Anzahl geteilt wird. Allein proportionalskalierte Daten verfügen zudem über geometrische bzw. harmonische Mittelwerte, die etwa die zentrale Tendenz von Wachstumsraten bzw. Verhältnissen beschreiben können. Andere Lageparameter sind beispielsweise der Mindest- und der Höchstwert, die bereits einen Eindruck davon vermitteln, wie weit die Daten verteilt (gestreut) sind. Solche Lageparameter reagieren aber sehr empfindlich auf Extremwerte (Ausreißer), so dass die beschreibende Statistik eigens besondere Streuungsparameter bereitstellt.

Streuungsparameter geben eine Antwort auf die Frage: Wie dicht um die Mitte der Daten herum liegen die Werte? Da nominalskalierte Daten keinerlei Ordnung und demnach keine Mitte aufweisen, lässt sich für sie auch kein Streuungsparameter errechnen.[106] Für mindestens ordinalskalierte Daten wird neben dem bereits erwähnten Abstand zwischen Mindest- und Höchstwert (Spannweite) vor allem der Abstand verwendet, innerhalb dessen die mittleren 50 % der Werte liegen (Quartilsabstand), also die Spannweite dessen, was übrigbleibt, wenn man jeweils das höchste und niedrigste Viertel der Werte abschneidet.[107] Für mindestens intervallskalierte Daten hingegen dient die Varianz als grundlegender Streuungsparameter. Varianz ist die mittlere quadrierte Mittelwertsabweichung, die sich ergibt, wenn für jeden einzelnen Datenpunkt sein Abstand zum Mittelwert errechnet, diese Differenz quadriert, für alle Werte aufsummiert und die Summe anschließend durch die Zahl der Werte geteilt wird.[108] Da die Varianz durch den Quadrierungsschritt eine andere Einheit aufweist als die ursprünglichen Daten, hat sich als leichter handhabbares Streuungsmaß die Standard-

---

[106] Eine ähnliche Funktion erfüllt aber der Simpson-Diversitätsindex, vgl. *Diekmann*, Sozialforschung 2012, 687.

[107] Vgl. *Krämer*, Gebrauchsanweisung 2010, 50.

[108] Die Quadrierung vermeidet nicht nur negative Abstandswerte – wie man oft liest (z.B. *Leary*, Methods 2008, 41) –, sondern beruht in erster Linie auf einer Eigenschaft des Mittelwerts: Mathematisch lässt sich zeigen, dass kein anderer Parameter eine geringere Summe *quadrierter* Abstände zu allen Werten einer Datenmenge hat.

abweichung etabliert, die nichts anderes ist als die Wurzel aus der Varianz. Ganz gleich wie eine Datenmenge aussieht, liegen mindestens 75 % der Daten innerhalb eines Abstands von zwei Standardabweichungen jeweils über- und unterhalb des Mittelwerts;[109] sind die Daten maßgeblich durch Zufallseinflüsse entstanden (wie Börsenkurse, die Länge von Vereinssatzungen oder das Alter von Kommanditisten), liegen sogar mehr als 95 % der Daten innerhalb dieses Abstands, davon ungefähr zwei Drittel innerhalb je einer Standardabweichung über- und unterhalb des Mittelwerts.[110]

Zusammenhangsparameter schließlich widmen sich der Frage: Wie gleichmäßig variieren mehrere Merkmale? Je nach Skalenniveau der Variablen für die beiden Merkmale sind zahlreiche Zusammenhangsparameter in Gebrauch.[111] Der bekannteste und für mindestens intervallskalierte Variablen üblichste ist der *Korrelationskoeffizient*. Er ist definiert als das Produkt der durchschnittlichen Mittelwertsabweichungen (Kovarianz) zweier Variablen, geteilt durch das Produkt ihrer Standardabweichungen. Praktisch bedeutet das, dass der Korrelationskoeffizient dimensionslos (ohne Maßeinheit) den Grad des linearen Zusammenhangs beziffert: Er ist genau dann +1, wenn jede Erhöhung der einen Variable um Betrag x einhergeht mit der Erhöhung der anderen Variable um Betrag y (der mit x identisch sein kann, aber nicht muss); dementsprechend ist er -1, wenn jede Erhöhung der einen mit einer gleichmäßigen Verringerung der anderen Variable einhergeht, und er ist 0, wenn keine der beiden Variablen eine *lineare* Beziehung zur anderen aufweist. Diese einseitige Betrachtung kann man durch eine beidseitige komplementieren, indem man den Korrelationskoeffizienten quadriert. Dann erhält man das *Bestimmtheitsmaß*, das angibt, welchen Prozentsatz ihrer Werteschwankung beide Variablen teilen, oder anders ausgedrückt: Den Anteil der Werteschwankung einer Variable, der durch die Werteschwankung der anderen *linear* erklärt werden kann. Das Bestimmtheitsmaß ist 0 wenn der Korrelationskoeffizient 0 ist (also die beiden Variablen miteinander nicht linear zusammenhängen) und 1 wenn der Korrelationskoeffizient +1 oder -1 (also eine Variable nur eine lineare Transformation der anderen) ist. Weil die Korrelation immer nur zwei Variablen bzw. Merkmale in Beziehung setzt, bedarf es für mehr als zwei Variablen (*multivariate* Beziehungen) einer allgemeineren Methode, die als

[109] Nach einer von *Tschebyschow* bewiesenen und nach ihm benannten Ungleichung liegt jeder Datenpunkt mit einer Wahrscheinlichkeit von höchstens $1/k^2$ mehr als k Standardabweichungen vom Mittelwert entfernt.

[110] Nach dem zentralen Grenzwertsatz bilden Summen von Zufallsvariablen eine sog. Normalverteilung (*de Moivre-Gauß*sche Glockenkurve), in der 68,27 % der Daten innerhalb je einer Standardabweichung und 95,45 % der Daten innerhalb je zweier Standardabweichungen über- und unterhalb des Mittelwerts liegen.

[111] Überblick bei *Beller*, Forschen lernen 2008, 72 f.; ausf. *Ellis*, Effect Sizes 2010, 11 f.

Regression bezeichnet wird (von lat. *re* und *gradi*, zurück gehen)[112] und die errechnet, welche Schwankungen der abhängigen Variable auf die Schwankungen mehrerer unabhängiger Variablen zurückgehen.[113] Dabei ist „zurückgehen“ nur in dem Sinn zu verstehen, in dem die linke Seite einer mathematischen Gleichung auf die rechte „zurückgeht“, nicht aber als empirischer Ursachenzusammenhang. Historisch kam die Regressions- vor der Korrelationsrechnung, daher wird der Korrelationskoeffizient bis heute mit „r“ (wie Regression) abgekürzt.[114]

Mit den von der beschreibenden Statistik ermittelten Parametern lassen sich Datenmengen natürlich nicht nur zusammenfassen, sondern auch vergleichen. Dabei sind vor allem zwei Vergleichsgrößen gebräuchlich: Mittelwertsunterschiede und Korrelationskoeffizienten.[115] Im Vergleich zweier *Grundgesamtheiten* ist die jeweilige Vergleichsgröße ein direktes Maß für den Unterschied oder Zusammenhang zwischen ihnen, also für die Größe des *Effekts*, den die Zugehörigkeit zu einer bestimmten Grundgesamtheit hat. Der Forscher hat dann nur noch zu klären, ob er diese Effektgröße für hinreichend bedeutsam hält oder nicht (dazu unten V.). Soweit allerdings zwei *Stichproben* verglichen werden, interessiert ja letztlich nicht die Effektgröße zwischen den Stichproben, sondern die Effektgröße zwischen den zugrundeliegenden Grundgesamtheiten. Zusätzlich ist dann also zu klären, ob der Stichprobeneffekt nicht nur auf Zufallsschwankungen bei der Stichprobenziehung beruht. Dazu dient die schließende Statistik.

2. Schließende (inferentielle) Statistik

Wie eingangs erläutert soll die schließende Statistik von einer Stichprobe auf ihre Grundgesamtheit verallgemeinern. Das betrifft den ganz überwiegenden Teil der empirischen Forschung, da Vollerhebungen schon aus Zeit- und Kostengründen oft nicht möglich sind. Stichprobenerhebungen dagegen laufen stets Gefahr, besonders untypische Merkmalsträger zu beobachten. Allgemeiner noch: Zwei Stichproben sind praktisch nie identisch, sondern werden immer einen Mittelwertsunterschied aufweisen. Wie lässt sich dieser Auswahleffekt vom eigentlich relevanten Haupteffekt trennen?

---

[112] *Harper*, Etymology 2014, Stichwort “regress”; Dudenredaktion, Fremdwörterbuch 2000, 1146 („Regress“); zum Ursprung des Begriffs vgl. *Salsburg*, Statistics 2001, 12 f., 44.

[113] *Beller*, Forschen lernen 2008, 72; *Weigand*, Statistik 2009, 111 ff. (Kap. 7); *Petersen/Goerg* in: Towfigh/Petersen, Methoden 2010, 201, 235 ff.; frühes juristisches Anwendungsbeispiel bei *Kötz/Schäfer*, AcP 1989, 501.

[114] *Stanton*, J Stat Edu 2001, unter 2.; *Diekmann*, Sozialforschung 2012, 360 Fn. 11: „Die Bezeichnung Regressionsanalyse ist ziemlich irreführend, ist heute aber nicht mehr wegzudenken.“

[115] Beide gehören zu je einer Familie von Effektgrößen (d- bzw. r-Familie), dazu näher unten bei Fn. 228.

Das gelingt vor allem durch eine zufällige Stichprobenziehung, die systematische Abweichungen der Stichprobe von der Grundgesamtheit verhindert.[116] Solche Stichproben können auch als *repräsentativ* bezeichnet werden, doch vermeide ich diesen Begriff, weil er umgangssprachlich oft so verstanden wird, dass die Stichprobe ihrer Grundgesamtheit ähnlich *ist*.[117] Stattdessen genügt es aber für die schließende Statistik, dass keine Abweichungen der Stichprobe von ihrer Grundgesamtheit *zu erwarten sind*. Die einzige Zahlenfolge, von der man erwarten darf, dass sie zu keiner anderen Zahlenfolge systematische Unterschiede (oder systematische Ähnlichkeiten) aufweist, ist die Zufallsfolge.[118] Dementsprechend ist die *einfache Zufallsziehung* von allen entwickelten Methoden der Stichprobenziehung die meistverwendete.[119] Sie minimiert das Risiko, dass besonders untypische Werte betrachtet werden, und ermöglicht eine genaue Bezifferung der Wahrscheinlichkeit, mit der ein bestimmter Mittelwertsunterschied zu erwarten ist, wenn tatsächlich kein systematischer Unterschied zwischen den Stichproben besteht.[120] Das ist das Grundanliegen der schließenden Statistik, und der Grund dafür, dass quantitative empirische Forschung fast immer mit Zufallsstichproben arbeitet.

Damit bleibt die Frage, *wie* die schließende Statistik vorgeht. Nur ein grundlegendes Verständnis dieser Methodik ermöglicht eine verständige Rezeption empirischer Forschung. Leider sind Einführungen in statistische Methoden im rechtswissenschaftlichen Schrifttum rar gesät,[121] was auch

---

[116] Anders scheinbar *Diekmann*, Sozialforschung 2012, 195: „Für hypothesenprüfende Untersuchungen können aber durchaus auch »willkürliche« Stichproben herangezogen werden." Das gilt zwar im Verhältnis zur Bevölkerungsgesamtheit; „Grundgesamtheit" i.S. der statistischen Auswertung sind jedoch die Studienteilnehmer, deren Zulosung in experimentelle Versuchsbedingungen letztlich doch eine zufällige Stichprobenziehung darstellt.

[117] Vgl. auch *Quatember*, Statistik ohne Formeln 2011, 121; *Bortz/Döring*, Forschungsmethoden 2006, 397.

[118] Oder umgekehrt: Prozesse, die neue Zahlenfolgen ohne *systematische* Unterschiede zu bestehenden erzeugen, heißen zufällig. Das ist unvermeidlich tautologisch (denn „systematisch" sind nur Unterschiede, die nicht zufällig sind), weil „Zufall" gerade der *Verzicht* auf Erklärungen ist, also mathematisch undefinierbar und rein empirisch existent.

[119] Zu anderen etwa *Beller*, Forschen lernen 2008, 87 f., 124 ff.; *Lercher* in: Barta u.a., Rechtstatsachenforschung 2009, 205, 213 f.; *Bortz/Döring*, Forschungsmethoden 2006, 394 ff. (Kap. 7); *Zeisel/Kaye*, Figures 1997, 101 ff. (Kap. 7); *Shadish/Cook/Campbell*, Causal Inference 2002, 342 ff.; *Chui* in: McConville/Chui, Methods 2007, 46, 55 f.

[120] Wie eingangs erläutert (oben IV. vor 1. a.E.), ist diese Wahrscheinlichkeit genau die, mit der die verglichenen Beobachtungen zur selben hypothetischen Grundgesamtheit gehören.

[121] Rudimentär *Hamann* in: Schröder u.a., FS Assenmacher 2012, 307; *Steckel-Berger* in: Barta u.a., Rechtstatsachenforschung 2009, 185; *Petersen/Goerg* in: Towfigh/Petersen, Methoden 2010, 201, 212 ff.; ausf. *Lawless/Robbennolt/Ulen*, Methods 2010, 189 ff. (Kap. 8–12); anspruchsvoll *Hagen*, Statistik für Juristen 2005; sehr forensisch orientiert *Zeisel/Kaye*, Figures 1997; *Finkelstein*, Basic Concepts 2009.

damit zusammenhängen mag, dass schließende Statistik als außerordentlich kompliziert gilt. Diese Wahrnehmung beruht allerdings weitgehend auf einem Missverständnis: Schließende Statistik ist nicht kompliziert, sondern *unübersichtlich*. „Egal wie unerfahren man in der Statistik ist, kommt man mit gesundem Menschenverstand und einem gründlichen Blick auf Tabellen, Graphen und Methodenberichte schon sehr weit.“[122] Schon das gründliche und strukturierte Herangehen also, das Juristen allgemein auszeichnet, ermöglicht ein solides Grundverständnis der Statistik, ohne dass es besonderer mathematischer Fähigkeiten bedürfte.[123] Dieses Grundverständnis sollen die weiteren Abschnitte vermitteln.

a) Schließende Statistik als angewandte Rhetorik

Statistik gehört nominell zur stochastischen Mathematik, und schon dieser Status verleiht ihr die Aura des Unabweichlichen, des Unangreifbaren und endgültig Wahren.[124] Umso mehr in der Vorstellung von Juristen, denen schon im Studium eingeredet wird, Rechnen sei nicht ihr Metier; diese „eher bedenkliche“ Umdeutung einer überlieferten Verfahrensregel (*iudex non calculat*)[125] verdeckt allerdings, dass „juristische und mathematische Begabungen gerade nicht ‚auf Kriegsfuß‘ stehen“, denn „hier wie dort wird die Fähigkeit zu analytischem Denken und logischer Schlussfolgerung in besonderer Weise vorausgesetzt“.[126] Zudem spielen mathematische Feinheiten in der angewandten Statistik eine ebenso untergeordnete Nebenrolle wie philosophische Feinheiten in der Rechtsdogmatik.

Dass Statistik oft die Sprache der Mathematik spricht, darf nicht darüber hinwegtäuschen, dass ihr Wesen demjenigen der Rechtswissenschaft deutlich ähnlicher ist als dem Wesen der analytischen Mathematik. Während etwa Arithmetik und Geometrie kraft *deduktiver* Logik aus einfachen Axiomen n-dimensionale Hyperwürfel konstruieren können,[127] die sich jeder menschlichen Vorstellungskraft entziehen, geht die Statistik *induktiv* von

---

[122] *Lempert*, J Emp L Stud 2010, 907, 926.

[123] *Epstein/Martin* in: Cane/Kritzer, Hdb Empirical 2010, 901, 913: “Conceptually, statistical inference is not all that hard to understand; actually we confront such inferences almost every day.”; *Lawless/Robbennolt/Ulen*, Methods 2010, 189: “It took mathematical genius to discover these techniques, but it does not take a genius to understand them.”

[124] Ähnl. *Abelson*, Principled Argument 1995, 7: “statistical calculations carry an aura of numerical exactitude”.

[125] Zur eigentlichen Bedeutung der Sentenz (heute § 319 I Var. 2 ZPO) vgl. *Liebs*, Rechtsregeln 2007, Nr. 150.

[126] *Hamann* in: Schröder u.a., FS Assenmacher 2012, 307, 308; vgl. auch *Hagen*, Statistik für Juristen 2005, 47; *Lawless/Robbennolt/Ulen*, Methods 2010, xix.

[127] Die Einsicht verdanke ich *Martin Menz*. Jüngst dazu *Hayes*, SdW 2012, 8/52: Mit zunehmendem *n* strebt das Würfelvolumen gegen unendlich, das Volumen der *größten* darin unterzubringenden Kugel jedoch gegen Null.

realen Phänomenen aus,[128] weil sie dem induktiven Grundanliegen der empirischen Forschung verpflichtet ist (ähnlich der Architektendevise *form follows function*).[129] In dieser praktischen Anwendung auf reale Sachfragen ähnelt die schließende Statistik der Rechtsdogmatik.[130] Daher überrascht es nicht, dass der „ideale Statistiker" in einer Kombination aus einem guten Geschichtenerzähler, einem guten Detektiv und einem guten Juristen gesehen wird.[131] Kaum eine andere Wissenschaftsdisziplin ist so von juristischen Metaphern durchsetzt wie die Statistik.[132]

Schließende Statistik strebt nicht nach sicherer Erkenntnis, sondern quantifiziert nur die im Erkenntnisprozess verbliebene Unsicherheit. Dadurch macht sie empirische Induktionsschritte zwischenmenschlich nachvollziehbar und diskursfähig, verfolgt also ein im Grunde *rhetorisches* Anliegen; sie lässt sich deshalb „als prinzipiengeleitete Argumentation" beschreiben.[133] Das bedingt, dass Statistik kommunikationsoffen sein muss, dass also hinter jeder statistischen Anwendung ein intuitiv verständliches Grundkonzept steht, über das auch der Laie grundsätzlich ein Plausibilitätsurteil fällen kann. Mit anderen Worten bezweckt Statistik intersubjektiv nachvollziehbare Entscheidungen (wie die Rechtsdogmatik), ist aber von objektiven Entscheidungen (nach Art der analytischen Mathematik) weit entfernt.[134] Dementsprechend toben in statistischen Fakultäten bisweilen heftige Auseinandersetzungen um die „richtige" Methode zur Quantifizie-

---

[128] Vgl. aber *Salsburg*, Statistics 2001, 31: Der Statistik-Pionier *Fisher* verwendete multidimensionale Geometrie, die für seine Zeitgenossen zu kompliziert war – und doch: "it is not too difficult to understand his conclusions." (44).

[129] Oder nach *van Belle*, Rules of Thumb 2008, 171: "analysis follows design"; dazu noch below B.II..

[130] Wenngleich auch die Rechtsdogmatik deduktiv vorzugehen versucht – und bisweilen Rechtsfiguren entwickelt, „die mit Mitteln des bloßen Denkens nicht mehr nachvollzogen werden" können (*AG München*, NJW 1985, 1230 zur Ersatzfähigkeit der fiktiven Mehrwertsteuer aus fiktiven Schäden).

[131] So *Abelson*, Principled Argument 1995, xiii, denn: "There are analogous features between the claims of a statistical analyst and a case presented by a lawyer" (xii); zur Detektiv-Metapher auch unten Fn. 146 sowie *Diekmann*, Sozialforschung 2012, 60 m.w.N.

[132] *Vickers*, Understand Statistics 2010, 140: "It is hard to think of any other area of science that is characterized with so many religious and legal metaphors."; weit verbreitetes Beispiel oben bei Fn. 80.

[133] So der Titel des Lehrbuchs von *Abelson*, Principled Argument 1995; ähnlich beschreibt *McCloskey*, Am Econ Rev 1985, 201; *McCloskey*, Rhetoric 1998, 112 ff. statistische Tests als immanent rhetorisch, und *Gorard*, Brit J Sociol Edu 2006, 67, 71 kritisiert: "researchers are relying on the false rhetoric of apparently precise probabilities, while abdicating their responsibility for making judgements about the value of their results."

[134] Als Indiz dafür nennt *Krämer*, Denkste! 2011, 214 m.w.N., dass selbst den größten Denkern in der Statistik viel häufiger gravierende Irrtümer unterlaufen als in der „reinen Mathematik"; vgl. aber *Salsburg*, Statistics 2001: "mathematical research is always filled with mistakes. Very few mathematicians can work alone."

rung empirischer Unsicherheit.[135] Rituale, Schwellenwerte und Testverfahren werden entwickelt, angegriffen und verworfen. Akademische Schulen entstehen und verschanzen sich im Grabenkampf, Wortgefechte werden mit scharfem, emotional aufgeladenen und fast ins Komische übersteigerten Eifer ausgetragen.[136]

Diese für das Bild der Mathematik völlig untypische subjektive Prägung liegt sicher auch daran, dass die moderne Statistik als Disziplin erst seit dem 20. Jahrhundert existiert und dass selbst ihre frühesten Wurzeln kaum weiter als bis in den Hochbarock zurückreichen. Daher überrascht es nicht, dass sogar grundlegende Methodenfragen noch ungeklärt sind. Andererseits werden niemals *alle* Methodenfragen zu klären sein, denn vielen vermeintlich technischen Streitfragen liegen divergierende philosophische Weltanschauungen zugrunde,[137] angefangen bei der Frage: Gibt es objektive Erkenntnis, oder ist Realität stets subjektive Konstruktion? Statistik ist also alles andere als zeitlos universelle Wahrheitssuche – sie ist eine vor allem durch soziale Normen und Konventionen geprägte rhetorische Übung. Es dürfte kaum einen Wissenschaftszweig geben, zu dem die Rechtsdogmatik konzeptionell ähnlich leicht Anschluss finden kann.[138]

b) Bayesianische und frequentistische Statistik

Ihre rhetorische Natur führt dazu, dass es verschiedene Schulen der Statistik gibt – und eine herrschende. Dass trotzdem sogar grundlegende konzeptionelle Unterschiede zwischen diesen Schulen in den meisten Statistiklehrbüchern ausgeblendet werden, wurde zwar zu Recht kritisiert,[139] ist hier aber kaum nachzuholen. Auch die folgende Darstellung wird sich also an

---

[135] *Krämer*, Denkste! 2011, 216 nennt das „heiße Glaubenskämpfe", ähnlich gar der Theologie; vgl. auch oben Fn. 132 sowie *Abelson*, Principled Argument 1995, xii.

[136] Paradebeispiel ist der bei *Lenhard*, Brit J Phil Sci 2006, 69 zusammengefasste Schlagabtausch zwischen *Fisher* und *Neyman* in der ersten Hälfte des 20. Jahrhunderts; ähnlich erbitterte Kontroversen gibt es heute etwa über den Unterschied zwischen substantieller und statistischer Signifikanz (unten bei Fn. 276), zwischen frequentistischen und bayesianischen Schulen der Statistik (dazu gleich b) oder zwischen klassischen linearen und modernen nicht-linearen (chaostheoretischen) Ansätzen, dazu im gesellschaftsrechtlichen Kontext *McKelvey/Andriani*, Strat Org 2005, 219.

[137] Ähnl. *Nickerson*, Psy Meth 2000, 241, 290: "lack of an accurate historical perspective and understanding of the complexity and sometimes controversial philosophical foundations of various approaches to statistical inference may go a long way toward explaining the apparent ease with which statistical tests are misused and misinterpreted."

[138] Krit. noch *Nußbaum*, AcP 1955, 453, 473: „Von allen Sozialwissenschaften eignet sich wahrscheinlich keine weniger für statistische Methoden als das Recht."; pragmatisch gelassen *Eisenberg*, J Am Stat Assoc 2000, 665, 668: "there is a need for advances in inferences involving small samples. Legal data are often categorical, and models often suffer from problems of endogeneity".

der herrschenden Übung (der sog. frequentistischen Statistik) orientieren – einem aus mancherlei Missverständnis geborenen und dementsprechend oft missverstandenen Hybrid aus eigentlich unvereinbaren Prinzipien.[140] Zumindest eine alternative Konzeption sei allerdings kurz erwähnt, weil sie aktuell enorme Aufmerksamkeit erfährt und in den nächsten Jahren noch wichtiger werden dürfte: die nach Thomas Bayes (1701–1761) benannte Statistik.[141] Ein Beispiel soll die unterschiedlichen Herangehensweisen von bayesianischer und frequentistischer Statistik illustrieren:[142]

Nehmen wir der Einfachheit halber an, es gäbe in einer Region Deutschlands exakt acht Unternehmen, davon vier Automobilhersteller und vier Einzelhandelskonzerne. Angenommen wir wissen auch, dass die Hälfte aller Automobilhersteller und drei Viertel aller Einzelhandelskonzerne liquide und solvent sind – ohne dass wir im Einzelnen wüssten, welche Unternehmen genau. Nun erhält der Staatsanwalt einen Hinweis von einem Wirtschaftsprüfer, der ein Unternehmen zufällig ausgewählt, überprüft und festgestellt hat, dass es längst insolvent ist und die Voraussetzungen des § 15a IV InsO zweifelsfrei vorliegen. Mehr weiß der Staatsanwalt nicht. Was soll er tun?

Auf diese Frage gibt es keine logisch zwingende Antwort, denn dem Staatsanwalt fehlt eine entscheidende Information: Hat der Wirtschaftsprüfer aus *einer* der beiden Branchen zufällig ausgewählt, oder aus *beiden*? Ohne diese Information ist das Problem nicht wohldefiniert und ein Wahrscheinlichkeitsurteil darüber, aus *welcher* Branche der Missetäter stammt,

---

[139] Etwa *Hubbard/Bayarri*, Am Stat'n 2003, 171: "Modern textbooks on statistical analysis […] typically present the subject matter as if it were gospel: a single, unified, uncontroversial means of statistical inference."; *Gigerenzer*, J Soc-Econ 2004, 587: "Textbooks […] almost never teach the statistical toolbox, which contains tools such as descriptive statistics, Tukey's exploratory methods, Bayesian statistics, Neyman–Pearson decision theory and Wald's sequential analysis."

[140] *Hubbard/Bayarri*, Am Stat'n 2003, 171: "essentially an anonymous hybrid consisting of the marriage of the ideas developed by Ronald Fisher on the one hand, and Jerzy Neyman and Egon Pearson on the other [vgl. oben Fn. 136]. It is a marriage of convenience that neither party would have condoned, for there are important philosophical and methodological differences between them"; *Blume/Peipert*, J Am Assoc Gyn Lap 2003, 439, 441; näher *Gigerenzer*, J Soc-Econ 2004, 587, 588, der die herrschende Übung als mechanisch und unreflektiert praktiziertes Ritual brandmarkt; zust. *Krämer*, KZfSS Sonderheft 2004, 51, 58 m.w.N.

[141] Vgl. aber *Salsburg*, Statistics 2001, 130: "there are a number of different methods of analysis and at least two different philosophical foundations […] as if entirely different ideas have been given the same label – Bayesian."; hier geht es um den zweiten dieser beiden Ansätze, den *Salsburg* "Personal Probability" (133 ff.) nennt.

[142] Frei adaptiert aus *Senn*, Appl Clin Trials 2003, 8/35; Bsp. aus der Rspr. bei *Zeisel/Kaye*, Figures 1997, 98.

schlicht nicht möglich.[143] Die zwei verschiedenen statistischen Ansätze umgehen dieses Problem auf verschiedene Weise:

Die bayesianische Statistik öffnet sich den subjektiven Überzeugungen des Staatsanwalts. Wenn er wirklich gar nichts weiß, nimmt er vielleicht an, der Wirtschaftsprüfer habe auf beide Branchen mit gleicher Wahrscheinlichkeit Zugriff. Dann könnte er insgesamt drei Insolvenzen aufdecken, davon zwei in der Automobilbranche, also wird der Staatsanwalt mit einer Wahrscheinlichkeit von zwei Drittel in der Automobilbranche fündig. Vielleicht hat er aber auch einen Anhaltspunkt dafür, warum der Wirtschaftsprüfer eher die eine als die andere Branche kennt. In der bayesianischen Statistik geht dieses subjektive a-priori-Vorwissen (engl. *prior*) gemeinsam mit der empirischen Beobachtung in eine a-posteriori-Überzeugung ein (*degree of belief*). Nach diesem Verständnis kann Statistik also immer nur subjektive Überzeugungen durch neue Beobachtungen aktualisieren, aber keine vom Interpreten losgelösten objektiven Wahrscheinlichkeiten liefern.[144] Ein zweiter Staatsanwalt mit anderem Vorwissen wird also eine andere Wahrscheinlichkeit errechnen.

Die frequentistische Statistik hingegen – die jünger ist als die bayesianische und genau diese Subjektivität eliminieren will – würde dem Staatsanwalt raten, schlicht jeden solchen Tipp (egal, woher er kommt) auf die weniger solvente Automobilbranche zu beziehen und nur zu errechnen, wie oft er daneben läge, wenn er konsequent so vorginge, also im Grenzfall unendlich oft (deshalb „frequentistisch"). Dann würde er in zwei von vier Fällen (also 50 %) vergeblich ermitteln. Dieses Vorgehen stützt sich ausschließlich auf gesicherte Informationen des Staatsanwalt und ist deshalb scheinbar objektiver.[145] Zugleich beantwortet es aber die falsche Frage: Der Staatsanwalt erfährt nur, wie wahrscheinlich ein Unternehmen insolvent ist, *wenn* es zur Automobilbranche gehört. Das ist etwas ganz anderes als die Wahrscheinlichkeit, dass ein Unternehmen zur Automobilbranche gehört, wenn es insolvent ist.[146]

---

[143] Amüsantere Illustrationen bei *Dubben/Beck-Bornholdt*, KZfSS Sonderheft 2004, 61.

[144] *Berger/Berry*, Am Sci 1988, 159, 165 m.w.N. in Fn. 8: Auch die sog. „objektive" bayesianische Methode beruhe nur auf der Unterstellung minimalen Vorwissens (also hier: gleiche Wahrscheinlichkeit für beide Branchen).

[145] Einfallstore für willkürliche Entscheidungen bietet auch die frequentistische Statistik, aber weniger offen als die bayesianische: "standard statistical methods turn out to be based on subjective input – input of a type that science should seek to avoid. […] it is hidden; few researchers realize how subjective standard methods really are.", *Berger/Berry*, Am Sci 1988, 159; vgl. auch *Sedlmeier/Renkewitz*, Forschungsmethoden 2008, 616.

[146] *Zeisel/Kaye*, Figures 1997 nennen das die "transposition fallacy"; anschaulich auch *Abelson*, Principled Argument 1995, 41 f.: "An analogy may be drawn with the situation faced by a detective solving a whodunit, who estimates that if the butler did it, there was only one chance in a hundred he could have made his escape unseen. Should the detective there-

Das Problem des Staatsanwalts ist eine triviale Illustration, aber auch ein Gleichnis für das Grundproblem der empirischen Forschung: Der Forscher („Staatsanwalt"), der mit seinen Instrumenten („Wirtschaftsprüfer") eine Beobachtung anstellt („Insolvenzreife"), will herausfinden, zu welcher Grundgesamtheit („Branche") die Beobachtung wahrscheinlich gehört, ohne *a priori* zu wissen, wie oft die eine oder die andere Grundgesamtheit in der Natur überhaupt zu beobachten sind (engl. *base rate*). Darüber kann er entweder spekulieren (bayesianischer Ansatz), oder er muss auf die Zugehörigkeit zu einer Grundgesamtheit daraus schließen, dass das Beobachtete in allen anderen Grundgesamtheiten zu selten vorkommt (frequentistischer Ansatz) – selbst wenn in der Natur vielleicht ausschließlich jene anderen Grundgesamtheiten auftreten.

Die beiden dargestellten Ansätze sind nur zwei von vielen möglichen statistischen Herangehensweisen, und sie ergänzen sich eher, als dass sie sich ausschließen.[147] Dass sich dennoch einer der beiden durchgesetzt hat, ist wiederum Ausdruck der rhetorischen Wurzeln der Statistik. Mittlerweile sind einige unvermeidliche Beschränkungen der frequentistischen Statistik zu Tage getreten, und Computer bieten ausreichend Rechenleistung, um die mathematisch anspruchsvollen bayesianischen Methoden durchzurechnen; daher gewinnen diese derzeit an Aufwind.[148] Völlig ablösen werden sie die frequentistische Statistik aber sicher ebensowenig wie andersherum.

c) Grundschema der frequentistischen Statistik

Gehen wir einstweilen von der herrschenden (frequentistischen) Methode aus, und betrachten deren Vorgehen näher. Wie bereits erwähnt, soll sie ermitteln, wie oft eine mit Hilfe der beschreibenden Statistik errechnete Effektgröße allein aus der Dynamik der Zufallsziehung resultiert, also *wenn* zwischen den Grundgesamtheiten tatsächlich kein Unterschied besteht. Das sagt nichts über die Fehlerwahrscheinlichkeit im konkreten Fall, macht aber intersubjektiv aushandelbar, welche Fehlerquote auf lange Sicht als hinnehmbar empfunden wird.

Damit bestimmt werden kann, wie oft zwischen zwei Stichproben ein Effekt zu beobachten ist, wenn in der Grundgesamtheit keiner vorliegt, muss vorher klar sein, um welchen Effekt es geht. Das ergibt die vor der

---

fore conclude there is only one chance in a hundred that the butler is the murderer? […] Suppose that the only other suspects are the maid and the chauffeur. If the maid did it, the chances that she would have escaped unseen were only one in a thousand. And for the chauffeur, even less. But somebody must have done it."

[147] *Howard/Maxwell/Fleming*, Psy Meth 2000, 315; *Sedlmeier/Renkewitz*, Forschungsmethoden 2008, 616.

[148] Knappe Einführung etwa bei *Sedlmeier/Renkewitz*, Forschungsmethoden 2008, 601 ff. (Kap. 19.2); *Bortz/Döring*, Forschungsmethoden 2006, 455 ff. (Kap. 7.2.5).

Datenerhebung aufgestellte Forschungshypothese, die auch Alternativhypothese genannt wird, weil sie sogleich einer sog. Nullhypothese gegenübergestellt wird.[149] Diese Nullhypothese ist das genaue logische Gegenteil der Alternativhypothese. Das heißt für jede der beiden Hypothesen, dass sie in allen Fällen zutrifft, in denen die andere Hypothese nicht zutrifft; ihre Wahrscheinlichkeiten ergeben also in der Summe 100 %.[150] Welche Wahrscheinlichkeiten das sind, werden wir – wie oben dargestellt – nie erfahren. (Dass Lehrbücher regelmäßig das Gegenteil behaupten,[151] beruht „hoffentlich" nur auf sprachlicher Sorglosigkeit.) Stattdessen nehmen wir probehalber an, dass die Nullhypothese zutrifft, fragen also letztlich, ob die Daten allein durch *Zufalls*schwankungen entstanden sein können. Dazu müssen wir errechnen, wie oft wir bei unendlich vielen Zufallsziehungen erwarten dürfen, die beobachteten oder noch extremere Daten zu erhalten; dafür gibt es zwei Arten von Verfahren: einfache *nicht-parametrische*, die die einzelnen Messwerte verwenden, und anspruchsvollere *parametrische*, die Lage- und Streuungsparameter in eine standardisierte *Testgröße* übersetzen,[152] deren Wahrscheinlichkeitsverteilung bekannt ist, weil sie sich bei unendlich vielen Zufallsziehungen einer bestimmten mathematischen Funktion annähert.[153]

Welche Testgröße verwendet wird, hängt allein davon ab, welcher Test zum Einsatz kommt, und das wiederum wird in erster Linie durch das Skalenniveau der Messgrößen bestimmt. Für jede Kombination von Skalenniveaus gibt es speziell entwickelte Tests.[154] Darüber hinaus gibt es viele auf diesen Grundformen aufbauende Tests, die bestimmte Zusatzannahmen treffen, um die Auswertung zu erleichtern oder bekannte Probleme zu umgehen.[155] Diese Tests haben exotische und bisweilen furchteinflößende Namen – wie „Kruskal-Wallis-Rangvarianzanalyse", „Pearsons χ^2-Anpassungstest" und „tetrachorische Korrelation" –, die sicherlich dazu beitra-

---

[149] Das Verfahren wird deshalb auch als *null hypothesis significance testing* (NHST) bezeichnet.

[150] Oder noch anders: Der Raum aller möglichen Wahrheiten wird in zwei disjunkte Mengen aufgeteilt.

[151] Z.B. *Leary*, Methods 2008, 249; *Röhl*, Rechtssoziologie 1987, 115.

[152] Testgrößen werden „verwirrenderweise häufig auch als Statistik (t-Statistik, F-Statistik, etc.) bezeichnet.", *Hagen*, Statistik für Juristen 2005, 139.

[153] Bspw. bilden die Mittelwerte wiederholt (Faustregel: mindestens 30 Mal) gezogener Stichproben aus mindestens intervallskalierten Daten eine sog. Normalverteilung (vgl. Fn. 110), die sich in eine standardisierte z-Verteilung übersetzen lässt – der daraus ablesbare z-Wert ist eine der wichtigsten Testgrößen. Für weniger als 30 Stichproben wird die z-Verteilung angepasst und heißt dann t(x)-Verteilung; x+1 bezeichnet die genaue Stichprobenzahl.

[154] Eine eingängige Systematik der wichtigsten Tests, samt Erläuterungen und Beispielen, bietet das Hochschulrechenzentrum der UCLA unter www.ats.ucla.edu/stat/stata/whatstat.

[155] Erläuterung und Entscheidungsbaum bei *Vorberg/Blankenberger*, PsR 1999, 157.

gen, Berührungsängste juristischer Rezipienten zu schüren. Dabei ist das juristische Handwerkszeug in zweierlei Hinsicht durchaus vergleichbar: Es verwendet erstens ähnlich bedrohliches Fachlatein (z.B. „actio pro socio", „materielle Unterkapitalisierung" oder „Insolvenzverschleppungshaftung") und bildet zweitens einen ebenso flexiblen Werkzeugkasten – manche Rechtsfiguren werden häufig gebraucht, andere so gut wie nie, niemand muss *alle* kennen oder gar ihre Voraussetzungen und Folgen auswendig wissen, und wenn keine der bisherigen passt, lassen sich immer neue entwickeln. Ganz genauso verhält es sich mit statistischen Tests. Es erleichtert die Anwendung solcher Tests, mit möglichst vielen schon gearbeitet zu haben, aber für das Verständnis eines bestimmten Tests oder die Auswahl eines Tests unter gegebenen Voraussetzungen ist das nicht zwingend erforderlich. Jedenfalls der Rezipient empirischer Forschung benötigt also kein tiefergehendes Wissen als das Bewusstsein, dass Tests und Testgrößen aufgrund praktischer Erfordernisse entwickelt und anhand ihrer empirischen Bewährung[156] nach den Umständen des Einzelfalls ausgewählt werden.

Indem wir der Wahrscheinlichkeitsverteilung einer Testgröße entnehmen, mit welcher relativen Häufigkeit ein bestimmter Wert dieser Testgröße aus Zufallsprozessen entsteht, erfahren wir, wie oft die Parameter, aus denen wir die Testgröße errechnet haben, rein zufällig zustande kommen. Für diese Wahrscheinlichkeit definieren wir einen Schwellenwert, das sog. Signifikanzniveau (üblicherweise 5 %),[157] und entscheiden wie folgt: Wenn in einem reinen Zufallsprozess, also bei unterstellter Wahrheit der Nullhypothese, in mindestens fünf von 100 Fällen die Daten zu beobachten gewesen wären, die wir beobachtet haben (oder extremere), weisen wir die Alternativhypothese zurück, da wir nicht ausschließen können, dass die Daten auch dann zustande kämen – durch die natürliche Streuung der Messwerte – wenn diese Hypothese nicht zutrifft. Andernfalls ist das Ergebnis *signifikant* und wir weisen die Nullhypothese zurück.

Wohlgemerkt bedeutet ein insignifikantes Ergebnis *nicht*, dass die zurückgewiesene Alternativhypothese falsch ist.[158] Ein insignifikantes Ergebnis bedeutet nur, dass uns die Stichprobenwerte nicht hinreichend überraschen, um den Glauben an die Nullhypothese aufzugeben (Falsifikation). Anders ausgedrückt: Aus der Stichprobe ist nicht mit hinreichender Sicherheit festzustellen, in welcher der verglichenen Grundgesamtheiten der Ver-

---

[156] *Vickers*, Understand Statistics 2010, 140 kennzeichnet Statistik deshalb gar als experimentelle Wissenschaft.

[157] Beachte aber *Gorard*, Brit J Sociol Edu 2006, 67, 70: "Standard limits […], such as 5%, have no mathematical or empirical relevance. They are arbitrary thresholds for decision-making."

[158] Prägnant *Aiken*, Edu Psy Meas 1994, 848, 856: "Absence of evidence is not evidence of absence."

gleichswert größer ausfällt.[159] Das *kann* daran liegen, dass beide gleich groß sind (der Effekt also exakt null ist), aber ebensogut kann es an der zu kleinen Stichprobe oder zu großen Datenstreuung liegen. Es bleibt also dabei: *Ob* oder wie wahrscheinlich die Alternativhypothese stimmt, wissen wir nicht, denn allen Berechnungen der frequentistischen Statistik liegt die Annahme zugrunde, dass die Nullhypothese wahr (die Alternativhypothese also falsch) ist. Das ist der Preis, den frequentistisch arbeitende Forscher dafür zahlen, miteinander überhaupt über Fehlerwahrscheinlichkeiten reden zu können. Es ist ein hoher Preis, denn das Signifikanztesten ist zwar „heute Standard quer durch alle Wissenschaften", aber auch „eine konstante Quelle von Irrtümern und Konfusion", weil oft verkannt wird, dass „eine signifikante Statistik für sich allein genommen überhaupt nichts" darüber aussagt, ob die „Ausgangshypothese falsch ist oder nicht, ob sie eklatant verletzt ist oder nur ein wenig oder ob die Abweichung von der Ausgangshypothese in irgendeiner Weise praktisch von Bedeutung ist".[160]

3. Erkundende (explorative) Statistik

Eine dritte Art statistischer Auswertung ist die erkundende Statistik. Dabei handelt es sich nicht um ein weiteres Methodenarsenal der Statistik, sondern um die Bezeichnung für einen bestimmten Einsatz der bereits dargestellten statistischen Methoden. Erkundende Statistik ist nicht mehr und nicht weniger als schließende Statistik ohne vorgängige Hypothese.[161] Diese Verwendung von Statistik ermöglicht es zwar, Regelmäßigkeiten in den Daten zu erkennen, erlaubt aber mangels vorgängiger Hypothese keine stochastisch zulässigen Aussagen über den Wahrscheinlichkeitsgrad der beobachteten Regelmäßigkeiten. Deshalb haben die Ergebnisse einer erkundenden Statistik ganz genau denselben Erkenntniswert wie weniger formalisierte anekdotische Evidenz (dazu oben I.1.): Sie sind wertvoll dabei, überprüfbare empirische Hypothesen zu bilden,[162] ermöglichen aber selbst keine systematische empirische Erkenntnis.

---

[159] *Abelson*, Principled Argument 1995, 10 Fn. 4 m.w.N.

[160] *Krämer*, Denkste! 2011, 189; nach *Dubben/Beck-Bornholdt*, KZfSS Sonderheft 2004, 61, 63 ist das der „häufigste und folgenreichste Irrtum der modernen internationalen medizinischen Forschung"; relativierend dagegen Nickerson, Psy Meth 2000, 241, 289: "Claims to the contrary notwithstanding, there is room for doubt as to whether [acquisition of knowledge] has been greatly impeded by the prevalence of such beliefs or by any of the many other shortcomings of NHST that have been ably identified by its critics."

[161] Vgl. *Hussy/Schreier/Echterhoff*, Forschungsmethoden 2010, 137.

[162] *Shadish/Cook/Campbell*, Causal Inference 2002, 31: "exploratory experiments that were unguided by formal theory and unexpected experimental discoveries […] have repeatedly been the source of great scientific advances."; vgl. auch *Epstein/King*, U Chi L Rev 2002, 1, 64; *Croson*, U Ill L Rev 2002, 921, 942.

Es gehört zu den großen Schwächen der empirischen Sozialforschung, dass sich fast nie abschätzen lässt, zu welchem Zeitpunkt eine bestimmte Hypothese gebildet wurde.[163] Oft ergeben sich während der Datenauswertung interessante Muster, deren anekdotische Natur nicht erkannt oder nicht respektiert wird. Stattdessen werden dann *post hoc* neue Hypothesen formuliert, die nicht etwa in einer zweiten Studie überprüft werden, sondern noch anhand derselben Daten, die ihre Formulierung inspiriert haben – ein eindeutiger Zirkelschluss. Dieses Vorgehen ist selten nachweisbar, führt aber zu einem scheinbar klaren „Ergebnis", das sich wesentlich leichter veröffentlichen lässt als Nullergebnisse (dazu unten bei Fn. 172).

V. Deutung

„Gibt es noch etwas zu sagen, nachdem alle Ergebnisse dargestellt sind? Häufig sogar sehr viel!"[164]

An die Auswertung der Daten schließt sich die Deutung der Ergebnisse an. Hier geht es darum, den Kreis zur Ausgangsfrage zu schließen, Schlussfolgerungen zu formulieren, Grenzen der Untersuchung zu erläutern, Verbesserungen und neue Fragen anzuregen. Den Ergebnissen muss hier *Bedeutung* entlockt werden.[165] Das umfasst in der Anwendungsforschung ohnehin, aber oft sogar in der Grundlagenforschung, mögliche normative Konsequenzen (*policy implications*). Der juristische Rezipient empirischer Forschung findet hier zwei mögliche Varianten:

Entweder hat eine empirische Disziplin gar kein normatives Erkenntnisinteresse oder wenigstens kein kohärentes Wertungsgebäude (wie große Teile der Psychologie[166]), dann beschränken sich Studien meist ohnehin auf Anregungen, wie sich die neuen Erkenntnisse bewerten lassen *könnten*. Juristische Rezipienten sind dann gefordert, ihrerseits eine normative Bewertung der Ergebnisse zu leisten und die neue Erkenntnis über das Sein (die Rechtswirklichkeit) auf den Stand der Erkenntnis über das Sollen (die Rechtsdogmatik) zu beziehen.

---

[163] *Abelson*, Principled Argument 1995, 58: "researchers are very inventive at concocting potential explanations of wrong-tailed results."

[164] *Beller*, Forschen lernen 2008, 161.

[165] *Ellis*, Effect Sizes 2010, 108: "Extracting meaning from research results is a challenge that many researchers avoid."; *Vickers*, Understand Statistics 2010, 100: "Many people seem to think that we statisticians spend most of our time doing calculations, but that is perhaps the least interesting thing we do. […] Statistics is more than just cutting and pasting from one software package to another. We have to think about what the numbers mean and the implications for our scientific question."

[166] Vgl. aber *Hussy/Schreier/Echterhoff*, Forschungsmethoden 2010, 18 zum „Verändern" durch Korrektur, Förderung oder Prävention als einem Forschungziel der Psychologie.

Oder die empirische Disziplin hat eine kohärente normative Theorie (wie die soziologische Rollentheorie oder die ökonomische Wohlfahrtstheorie), in deren Rahmen sie beurteilt, wie die neue Erkenntnis bewertet werden *sollte*. Dann muss der juristische Rezipient überprüfen, inwieweit diese normative Folgerung mit juristischen Wertungskategorien vereinbar ist, und etwaige Abweichungen wertend begründen.[167] In jedem Fall muss der juristische Rezipient die normative Ausdeutung der Ergebnisse als sein Hoheitsgebiet begreifen und nutzen; die empirische Forschung kann ihm nur Anregungen geben, aber keine Vorgaben machen.[168]

VI. Veröffentlichung

Sind die plangemäß erhobenen Daten ausgewertet und gedeutet, ist das empirische Forschungsprojekt noch nicht beendet. Im Gegenteil: Diejenige Phase, die oft am längsten dauert, hat noch nicht einmal begonnen. Dennoch widmen einschlägige Darstellungen des empirischen Forschungsablaufs der Veröffentlichung kaum Aufmerksamkeit. Wenn sie überhaupt darauf eingehen, geben sie in der Regel Hinweise für die Struktur und den Stil von Manuskripten.[169] Zumindest für den Rezipienten empirischer Forschung hingegen ist die Veröffentlichungsphase unter einem ganz anderen Gesichtspunkt von überragender Bedeutung: Sie ist *das* maßgebliche Nadelöhr der empirischen Forschung.

1. Qualitätskontrollen

In den meisten empirischen Disziplinen folgen Veröffentlichungen demselben Grundprinzip. Das wichtigste Medium sind Zeitschriften. Anders als bei Juristen sind Monographien von untergeordneter Bedeutung und erscheinen meist in Gestalt von Lehr- und Handbüchern oder als groß angelegte Forschungssynthese. Reguläre akademische Veröffentlichungen dagegen erfolgen fast ausschließlich in Zeitschriften. Das schließt auch Qualifi-

---

[167] Dazu ausf. *Engel* in: Engel u.a., Recht und Verhalten 2007, 363, 387 ff.; *Garrn*, Rationalität 1986, 117: „der rechtlich Entscheidende“ müsse „berücksichtigen, von welchem Vorverständnis sozialer Wirklichkeit die betreffende sozialempirische Untersuchung geleitet ist, und überlegen, ob dieses […] zu überzeugen vermag.“

[168] *Engel* in: Engel, Methodische Zugänge 1998, 11, 40; *Schneider/Teitelbaum*, Utah L Rev 2006, 53, 66: “lawyers may feel free – almost obliged – to integrate the empirical and the normative in a way many social scientists will not but somebody must.”; krit. *Ietswaart* in: Plett/Ziegert, Empirische Rechtsforschung 1984, 210, 213 zur “tendency to be afraid of seeing the researcher place the data clearly in a theoretical framework. [.] policy-makers [.] want to reserve the right to draw conclusions.”

[169] Etwa *Beller*, Forschen lernen 2008, 153 ff. (Kap. 7); *Bortz/Döring*, Forschungsmethoden 2006, 86 ff.; fundiert und ausf. auch *Epstein/Martin/Schneider*, Vand L Rev 2006, 1811 und *Epstein/Martin/Boyd*, Vand L Rev 2007, 801.

kationsarbeiten ein, z.B. kumulative Dissertationen und die anhaltende Veröffentlichungstätigkeit, die anstelle der bei Juristen üblichen Habilitation den Zugang zur akademischen Laufbahn eröffnet. Aufgrund dieser überragenden Rolle von Zeitschriftenveröffentlichungen haben empirische Disziplinen ein komplexes System der Qualitätskontrolle.

Zum einen werden Zeitschriften danach bewertet, welchen Einfluss sie auf den Diskurs haben: Für jede Zeitschrift wird laufend errechnet, wie oft ihre Inhalte durchschnittlich zitiert werden (*impact factor*), und zwar anhand verschiedener Errechnungsverfahren, die zahlreiche Korrekturfaktoren berücksichtigen, um aussagekräftige Werte zu erhalten und Manipulationen zu erschweren. Je höher dieser Einfluss der Zeitschrift, desto begehrter und höher angesehen ist eine Veröffentlichung darin. So entstehen für jede Disziplin und ihre Subdisziplinen Rangordnungen (*journal rankings*), an deren Spitze meist die disziplinübergreifenden Zeitschriften *Science* und *Nature* rangieren.[170]

Zum anderen praktizieren die meisten Zeitschriften die Kreuzbegutachtung (*peer review*):[171] Eingereichte Texte werden zunächst von einem Schriftleiter mit wissenschaftlicher Qualifikation gesichtet. Was er nicht von vornherein ablehnt, leitet der Schriftleiter an mindestens zwei ausgewählte Wissenschaftler mit einem Schwerpunkt im betreffenden Themengebiet weiter, die er (freiwillig und ehrenamtlich) um die Erstellung eines Gutachtens bittet. In der Regel werden die Manuskripte vorher anonymisiert und die Gutachter um eine von drei Empfehlungen gebeten: Ablehnung, Annahme oder Überarbeitung. Falls der Schriftleiter dieser Empfehlung folgt, kann der Autor im ersten Fall seinen Text bei einer anderen Zeitschrift einreichen (gleichzeitige Paralleleinreichungen sind meist untersagt), im zweiten Fall damit rechnen, dass sein Text zügig online erscheint und innerhalb von einigen Monaten bis Jahren abgedruckt wird, und im dritten Fall seinen Text überarbeiten und erneut einreichen (*revise and resubmit*); damit beginnt das Verfahren von neuem, nur dass am Ende oft nur noch eine Annahme oder Ablehnung in Frage kommt. Aufgrund der Arbeitslast der Schriftleiter und Gutachter, die diese Funktion *neben* ihrer wissenschaftlichen Forschung ausüben, streckt sich das Verfahren nicht selten über mehrere Jahre.

---

[170] Eine auf dem einflussreichsten Zitatenindex (Thomson ISI Journal Citation Report) beruhende Übersicht von Rangordnungen findet sich unter www.sciencegateway.org/rank; die bislang einzige Rangordnung deutscher Rechtszeitschriften (umfrage-, nicht einflussbasiert) bieten *Gröls/Gröls*, JZ 2009, Beil. H. 17, 32.

[171] *Engel*, JITE 2010, 199, 200: "These days, the prime disciplinary governance tool is peer review."; vgl. auch *Epstein/King*, U Chi L Rev 2002, 1, 48.

2. Strukturelle Engpässe

Die Qualitätskontrollen im Veröffentlichungsverfahren haben einige erhebliche Nachteile. Zunächst führt die Kreuzbegutachtung beinahe zwingend zu einer Verknappung der Veröffentlichungskanäle. Denn damit nicht der eigentliche Forschungsbetrieb zum Erliegen kommt, kann jeder Wissenschaftler nur eine begrenzte Anzahl von Begutachtungsaufträgen übernehmen, was wiederum die Anzahl der kreuzbegutachteten Zeitschriften beschränkt. Da Zeitschriften ohne Kreuzbegutachtung (oder etwa freie Onlineveröffentlichungen) kein Ansehen genießen und kaum wahrgenommen werden, stehen Forscher also im permanenten Wettbewerb um die knappen Veröffentlichungsgelegenheiten.

In diesem Wettbewerb ist zu beobachten, dass deutlich mehr Studien veröffentlicht werden, die einen vermuteten Effekt empirisch belegen als Studien, die einen Effekt vermuten, dafür aber keine signifikanten Ergebnisse finden („Nullergebnis").[172] Wenn die Autoren empirischer Studien nicht gerade hellseherische Fähigkeiten haben,[173] lässt diese Asymmetrie darauf schließen, dass Schriftleiter und Gutachter gezielt solche Texte zur Veröffentlichung auswählen, die signifikante Ergebnisse berichten, während sie Nullergebnisse überwiegend ablehnen.[174] Dieses Vorgehen wird – wenn überhaupt – damit zu rechtfertigen versucht, dass Nullergebnisse keine neuen Erkenntnisse beinhalten,[175] weil sie schlicht durch „lausige Studien" verursacht worden sein könnten.[176] Begründungen dieser Art sind allerdings derart fehlgeleitet, dass man sich wundern muss, sie sogar in Einfüh-

---

[172] Vgl. etwa *Leary*, Methods 2008, 22: "scientific journals are reluctant, if not completely unwilling, to publish studies that fail to obtain effects."; *Abelson*, Principled Argument 1995, 54 f.; das wurde in den letzten fünfzig Jahren wiederholt kritisiert: *Sterling*, J Am Stat Assoc 1959, 30; *Sterling/Rosenbaum/Weinkam*, Am Stat'n 1995, 108.

[173] So die polemische These von *Bones*, Persp Psy Sci 2012, 307.

[174] Nach einer Umfrage von *Coursol/Wagner*, Prof Psy Res Pract 1986, 136 (Tabelle 2) wurden 80 % der eingereichten Ergebnisse, aber nur 50 % der eingereichten Nullergebnisse veröffentlicht; fast identisch das Experiment von *Atkinson/Furlong/Wampold*, J Couns Psy 1982, 189 (82 % der Ergebnisse, 43 % der Nullergebnisse); auch im früheren Experiment von *Mahoney*, Cog Therapy Res 1977, 161, 168 (Tabelle 1) bewerteten Gutachter Studien mit Nullergebnissen schlechter (1,8 auf einer Skala 1–6) als identische Studien mit Ergebnissen (3,2) oder ganz ohne Resultat (3,4).

[175] *Sterling/Rosenbaum/Weinkam*, Am Stat'n 1995, 108 zitieren einen Schriftleiter mit dem Urteil: "the study was well documented. Unfortunately, the negative results translates [sic] into a minimal contribution to the field."

[176] So wörtlich *Leary*, Methods 2008, 22: "data may fail to support our research hypotheses for reasons that have nothing to do with the validity of a particular hypothesis. As a result, null findings are usually uninformative […] Was the hypothesis disconfirmed, or did we simply design a lousy study? Because we can never know for certain, journals generally will not publish studies that fail to obtain effects."

rungslehrbüchern zu lesen. Schließlich kann jedes Ergebnis – nicht nur Nullergebnisse – auf der schlechten Durchführung einer Studie beruhen. Die Gültigkeit einer Studie lässt sich nie an ihrem Ergebnis messen, sondern immer nur an ihrer Durchführung.[177] Deshalb wird die Studiendurchführung stets detailliert berichtet, ganz gleich ob am Ende ein Ergebnis oder Nullergebnis steht. Dass Nullergebnisse nicht veröffentlicht würden, weil sie auf einer schlechten Studiendurchführung beruhen *könnten*, ist demnach nur eine bequeme Scheinrechtfertigung.[178] Tatsächlich ist das Ungleichgewicht der Veröffentlichungslandschaft kaum zu rechtfertigen, sondern bestenfalls zu erklären: In der Realität gibt es weniger Zusammenhänge als Nichtzusammenhänge, deshalb finden Menschen erstere womöglich schlicht spannender als letztere.[179] Was immer die Ursache des Ungleichgewichts, es hat zwei verheerende Folgen:

Erstens werden Nullergebnisse oft gar nicht erst zur Veröffentlichung eingereicht.[180] Aus der verzerrten Auswahl durch die Veröffentlichungsorgane (*publication bias*) folgt also ein verzerrtes Veröffentlichungsbemühen durch die Autoren (*reporting bias*). Weil insignifikante Ergebnisse nicht eingereicht werden, sondern in der Schublade verschwinden (deshalb anschaulich *file drawer effect*),[181] entsteht eine kaum quantifizierbare Dunkelziffer an unveröffentlichten Studien, die trotz ähnlicher Ausgangsbedingungen wie ihre veröffentlichten Gegenstücke *keinen* Effekt finden. Wenn solche Ergebnisse der Forschungsgemeinschaft vorenthalten werden, wird die Bedeutung der veröffentlichten Effekte zwangsläufig überschätzt und die Forschung kann in ganz falsche Richtungen laufen:

> „Dieses Aussortieren von nicht ‚signifikanten' Studien durch Dritte lässt die Bedeutung der überlebenden Ergebnisse genauso unangemessen wachsen wie das Manipulieren der […] Studien selbst."[182]

---

[177] Vgl. auch *Zeisel/Kaye*, Figures 1997, 97: "A statistically significant fmding from a poorly designed study deserves no more credence than any other fmding from such a study."

[178] Im Gegenteil: Gerade *weil* Studien mit Nullergebnis kaum je veröffentlicht werden, sind sie vertrauenswürdiger, weil niemand Nullergebnisse vorsätzlich fälschen oder durch geschickte Auswertung herbeirechnen wird.

[179] Prägnant auf den Punkt gebracht von *Goldacre*, Guardian 2011: "Scientific journals can be as bad as newspapers in preferring eye-catching stories to negative findings."

[180] Nach einer Umfrage von *Coursol/Wagner*, Prof Psy Res Pract 1986, 136 (Tabelle 1) werden Nullergebnisse halb so oft (43 % statt 82 %), nach der Umfrage von *Greenwald*, Psy Bull 1975, 1, 4 sogar acht Mal seltener (5,9 % statt 49,4 %) zur Veröffentlichung eingereicht – und stattdessen vier Mal so oft sofort aufgegeben (28 % statt 6,6 %); vgl. auch *Abelson*, Principled Argument 1995, 54 f.: "Students may abandon dissertations because null hypotheses cannot be rejected."

[181] Begriff nach *Rosenthal*, Psy Bull 1979, 638; vgl. auch *Ellis*, Effect Sizes 2010, 117 f.

[182] *Krämer*, Denkste! 2011, 192; anschauliche Anekdote bei *Dubben/Beck-Bornholdt*, Fehlinformation 2011, 100 m.w.N.:„Diagoras von der Insel Milos […] wurde vor 2500 Jahren

Im Extremfall lässt diese Entwicklung sogar erwarten, dass „die meisten publizierten Forschungsergebnisse falsch sind".[183] Als Abhilfe haben einzelne Disziplinen Zeitschriften gegründet, die ausschließlich Nullergebnisse publizieren,[184] aber sämtlich *open-access*-Projekte sind und damit längst nicht so angesehen und beachtet wie kommerziell verlegte Zeitschriften.[185]

Zweitens erhalten Autoren einen erheblichen Anreiz, Hypothesen entweder von vornherein möglichst unscharf zu formulieren oder im Nachhinein (*post hoc*) an die Beobachtungen anzupassen (*HARKing*).[186] Beides widerspricht den oben dargestellten Grundlagen der schließenden Statistik und bedeutet bildlich gesprochen einen Zirkelschluss (vgl. oben IV.3. a.E.). In die Formeln der schließenden Statistik lassen sich zwar auch Ereignisse einsetzen, die bereits beobachtet wurden, aber alle so gewonnenen Wahrscheinlichkeitsschlüsse sind unsinnig, denn beobachtete Ereignisse haben per Definition eine Wahrscheinlichkeit von 100 %.[187] Wenn aber nicht klar ist, ob die statistischen Methoden und Hypothesen vorab festgelegt und damit aussagekräftig waren, oder ob sie nur erkundenden Wert haben, werden statistische Ergebnisse ganz generell entwertet.[188] Das führt dazu, dass viele einmal belegte Effekte später nicht mehr repliziert werden können, so dass die Effekte scheinbar über die Zeit abnehmen (*decline effect*), während sie in Wirklichkeit nie existierten.[189] Das zu verhindern könnten Datenbanken helfen, in denen Studien vorher detailliert registriert werden

---

angesichts der Danktafeln […], die gerettete Schiffbrüchige in einem Tempel aufgestellt hatten, von einem Priester gefragt, ob dies nicht ausreichende Evidenz für die Existenz der Götter sei. Worauf Diagoras entgegnet haben soll: ‚Und wo sind die Tafeln der Ertrunkenen?'".

[183] So der provokative Titel von *Ioannidis*, PLoS Med 2005, 696 (mit bayesianischer Begr.), dazu *Diekmann*, JBNST 2011, 628; ähnl. *Dubben/Beck-Bornholdt*, KZfSS Sonderheft 2004, 61, 74: „Ein Großteil dessen, was in ‚wissenschaftlichen' Journalen gedruckt wird, ist nicht valide."; ebenso *De Long/Lang*, J Polit Econ 1992, 1257 (modellanalytisch); krit. *Nickerson*, Psy Meth 2000, 241, 271.

[184] Bsp.: J of Articles in Support of the Null Hypothesis, J of Null Results, J of Negative Results Ecology & Evolutionary Biology, J of Pharmaceutical Negative Results oder J of Negative Results in BioMedicine; für die Psychologie www.psychfiledrawer.org (dazu *Carpenter*, Sci 2012, 1558; *Yong*, SdW 2013, 2/58).

[185] Auch zwei der zuletzt meistdiskutierten psychologischen Replikationen (vgl. nur *Yong*, Nature 2012, 298; *Yong*, SdW 2013, 2/58) landeten in einer *open-access*-Zeitschrift (*Doyen/Klein u.a.*, PLoS ONE 2012; *Ritchie/Wiseman/French*, PLoS ONE 2012), und erstere wurde prompt aufgrund dessen angegriffen (*Bargh*, Nothing in Their Heads 2012).

[186] Kurz für *Hypothesizing After the Results are Known*, näher *Kerr*, Pers Soc Psy Rev 1998, 196.

[187] Vgl. nur *Duckworth*, Teach Stat 2006, 84, 86.

[188] *Begg/Berlin*, J Royal Stat Soc A 1988, 419, 428: "This has the consequence, among many informed observers, of discrediting the value of the literature as a source of believable information and of the ability of conventional statistical methods to make reliable inferences."

[189] *Schooler*, Nature 2011, 437; *Lehrer*, New Yorker 2010.

müssen. Verbreitet sind solche Datenbanken bisher nur in der medizinischen Forschung;[190] sie hindern Forscher zwar nicht daran, so oft zu experimentieren, bis sie einen Effekt finden, machen die Zahl und Konfiguration der Versuche aber transparent.

3. Mangel an Replikationen

Neben der Bevorzugung signifikanter Ergebnisse führt der Veröffentlichungsprozess auch noch auf anderen Wegen zu Verzerrungen der Forschungslandschaft. Etwa veröffentlichen gerade die angesehensten Zeitschriften oft grundsätzlich keine Replikationsstudien und verweisen Autoren stattdessen auf eine von Hunderten weniger profilierter Zeitschriften.[191] Diese allerdings werden kaum wahrgenommen, so dass eine dort veröffentlichte erfolglose Replikation die Ursprungsstudie nicht ernstlich in Zweifel ziehen kann.[192] Zudem wird für erfolglose Replikationsversuche oft der Autor der Ausgangsstudie zum Gutachter bestellt. So wichtig die Gelegenheit für jenen Autor ist, sich zur mangelnden Replizierbarkeit seiner Ergebnisse äußern zu können, so gefährlich ist es auch, ihn als Gutachter zur Unterdrückung von Erkenntnissen zu ermächtigen, die seinen eigenen widersprechen und seinem Ansehen dadurch schaden könnten. Zumal hier regelmäßig ein strukturelles Machtgefälle herrscht: Studien, die interessant genug sind, um repliziert zu werden, befördern ihren Autor schnell in hohe Positionen. Autoren hingegen, die noch Studien replizieren statt eigene Forschungsrichtungen zu verfolgen, sind typischerweise oft in niedrigeren Positionen. Daher überrascht es nicht, dass viele Forscher Replikationen von vornherein vermeiden, um

> „es sich mit den Kolleginnen und Kollegen nicht [zu] verderben, die nicht selten äußerst empfindlich reagieren, wenn ihre ‚heißgeliebte' Hypothese, auf der gelegentlich auch noch gewaltige spekulative Gedankengebäude errichtet wurden, sich in einer Wiederholungsuntersuchung als äußerst mager erweist."[193]

Erkenntnisse aus Replikationsstudien sind im Veröffentlichungsprozess also systematisch benachteiligt, und doch sind gerade diese Erkenntnisse

---

[190] *De Angelis/Drazen u.a.*, New Engl J Med 2004, 1250: "The I[nternational]C[ommittee of]M[edical]J[ournal]E[ditors] member journals will require, as a condition of consideration for publication, registration in a public trials registry."; Evaluation z.B. bei *Mathieu/Boutron u.a.*, J Am Med Assoc 2009, 977.

[191] *Aldhous*, NewScientist 2011; *Crocker/Cooper*, Sci 2011, 1182: "studies that replicate (or fail to replicate) others' findings are almost impossible to publish in top [.] journals".

[192] Bsp. bei *Begg/Berlin*, J Royal Stat Soc A 1988, 419, 428 mit dem Fazit: "a highly visible publication can have a dramatic and prolonged impact on medical practice even if the results are demonstrated to be unreliable."

[193] *Diekmann*, Sozialforschung 2012, 70.

besonders wichtig:[194] Die gesamte frequentistische Statistik beruht auf einem Fehlerkalkül für den Fall, dass ein bestimmter Versuch beliebig oft wiederholt würde. Wenn solche Wiederholungen aber tatsächlich nie stattfinden oder nie berücksichtigt werden, verlieren Signifikanzaussagen ihre Glaubwürdigkeit.[195]

Nicht zuletzt deshalb ist es teilweise üblich, von Autoren für eine Veröffentlichung immer mindestens drei zusammenhängende Studien einzufordern, damit ein Mindestmaß an Replizierbarkeit des Haupteffekts sichergestellt ist.[196] Neuerdings rufen einzelne Zeitschriften sogar ganz ausdrücklich zur Einsendung von Replikationen auf und stellen deren Veröffentlichung konkret in Aussicht.[197] Bisher finden sich auch kaum Versuche, veröffentlichte Studien systematisch zu replizieren,[198] doch eine unlängst zusammengetretene Forschergruppe, die über eine offene Internetplattform kollaboriert, hat es sich sogar zum Ziel gesetzt, alle Studien zu replizieren, die im Jahr 2008 in drei der angesehensten psychologischen Zeitschriften erschienen sind.[199]

4. Subjektivität und Willkür

Schließlich ist auch Allzumenschliches dem akademischen Veröffentlichungsprozess nicht fremd. Schriftleiter können Texte für die Kreuzbegutachtung kaum effektiv anonymisieren, weil Autoren, die an gleichen Themen forschen, einander meist ohnehin kennen. Zudem sind in einigen Disziplinen Vorveröffentlichungen im Internet üblich (*preprints*),[200] um Ideen-

---

[194] Sehr illustrativ die Beiträge in *Verschiedene*, Psy'st 2012, 346 unter www.thepsychologist.org.uk/archive/archive_home.cfm?ArticleID=2059; vgl. auch Zitat bei *Nickerson*, Psy Meth 2000, 241, 284 m.w.N.: "An ounce of replication is worth a ton of inferential statistics".

[195] *Diekmann*, Sozialforschung 2012, 190: „Der Verzicht auf Replikationen kann [.] leicht zu einer Kumulation von Irrtümern führen."; *Salsburg*, Statistics 2001, 99 m.w.N. gegen "isolated significant results which [the researcher] does not know how to reproduce"; vgl. auch *Epstein/King*, U Chi L Rev 2002, 1, 38 ff. mit juristischen Bsp.

[196] *Abelson*, Principled Argument 1995, 133: "it is a standard journal editorial policy to require of authors that they report experiments in clusters of three or more interrelated studies, serving (among other things) to provide a modicum of replication."

[197] Beispielhaft *Alm*, Publ Fin Rev 2010, 4 mit gleichlautendem Aufruf in den sechs folgenden Ausgaben (zuletzt *Burman/Reed/Alm*, Publ Fin Rev 2011, 190) und rot hinterlegter (!) Verlinkung auf der Website http://pfr.sagepub.com; zu einem früheren Beispiel vgl. *Fuess*, Q J Bus Econ 1996, 3.

[198] Ausnahme etwa *Dewald/Thursby/Anderson*, Am Econ Rev 1986, 587, 600 mit dem Fazit: "Our results […] suggest that [.] errors may be quite common [… which] frustrates replication and prevents later researchers from building on earlier research."

[199] Online unter https://osf.io/ezcuj/, erläutert von *Open Science Collaboration*, Persp Psy Sci 2012, 657; *Carpenter*, Sci 2012, 1558; *Yong*, Nature 2012, 298; *Yong*, SdW 2013, 2/58.

[200] Z.B bei SSRN (papers.ssrn.com), RePEc IDEAS (ideas.repec.org/i/p.html) und arXiv (arxiv.org).

diebstahl zu verhindern[201] und Arbeiten frühzeitig auf Konferenzen und Seminaren zur Diskussion stellen zu können. Daher ermöglichen die in den Veröffentlichungsprozess eingebauten Qualitätskontrollen auf ihrer Kehrseite geradezu, dass auch persönliche Animositäten und subjektive Hintergedanken in die Veröffentlichungsentscheidung einfließen.[202]

Dass objektive, qualitätsgeleitete Veröffentlichungsentscheidungen sogar eher Ausnahme als Regel sein könnten, illustriert ein klassisches (und natürlich heftig umstrittenes) Experiment:[203] Bei zwölf hoch angesehenen psychologischen Zeitschriften wurde jeweils eine Studie zur Veröffentlichung eingereicht, die zwei bis drei Jahre zuvor *schon einmal* veröffentlicht worden war. (Nur Name und institutionelle Anbindung der betreffenden Autoren wurden durch fiktive Angaben ersetzt.) Gerade einmal drei von 38 beteiligten Schriftleitern und Gutachtern erkannten das Vollplagiat, von den übrigen neun Manuskripten wurden acht einhellig abgelehnt, teilweise aufgrund „ernsthafter methodischer Mängel". Dass lässt befürchten, dass Annahme oder Ablehnung einer Studie oft eher vom Zufall als von ihrer Qualität abhängen.[204]

Juristische Rezipienten müssen sich also bewusst sein, dass derjenige Teil der empirischen Forschung, der in kommerziell verlegten Zeitschriften dokumentiert ist, systematisch verzerrt ist. Dem zu begegnen versucht die siebte und letzte Phase des Forschungsprozesses.

VII. Synthese

In gewissen Abständen entstehen zwangsläufig Diskrepanzen im empirischen Erkenntnisrepertoire. Verschiedene Studien finden zu den gleichen Fragen unterschiedliche Ergebnisse – einerseits weil „gleiche" Fragen (und der Weg, sie zu beantworten) nie identisch sind und schon kleine Unterschiede große Folgen haben können, andererseits weil der Veröffentlichungswettbewerb Innovationen belohnt und Forscher zwecks Profilbildung versuchen, sich voneinander abzugrenzen. Früher oder später wird

---

[201] In Disziplinen ohne *preprint*-Verfahren sind durchaus Fälle bekannt, in denen Gutachter die Ablehnung eines Textes empfohlen und ihn später als eigenen ausgegeben haben. ("This happened many many times.", so Uwe Dirnagl im Vortrag "Good Scientific Practice" der 2nd Winter School "Ethics and Neuroscience" im BCCN Berlin am 19.2.2012; konkretes Beispiel etwa bei *Hamblin*, Brit Med J 1981, 1671).

[202] *Salsburg*, Statistics 2001, 192: "especially when there is some controversy associated with the subject, editors are tempted to publish that which is acceptable to the scientific community"; *Goldsmith/Vermeule*, U Chi L Rev 2002, 153, 162: "senior scholars favor like-minded scholarship and choke off the channels of intellectual change and development."

[203] *Peters/Ceci*, Behav Brain Sci 1982, 187.

[204] So auch das Ergebnis der Studien von *Cole/Cole/Simon*, Sci 1981, 881 und *Rothwell/Martyn*, Brain 2000, 1964.

der tatsächliche Erkenntnisstand also nicht mehr ohne Weiteres zu ermitteln sein, eine *Synthese* der Forschung wird erforderlich.

Forschungssynthese erfolgt kontinuierlich durch die Bildung und Verfeinerung von Theorien (dazu später 3.) und nimmt im Übrigen meist eine von zwei Formen an: die qualitative oder die quantitative. Damit sind nun – anders als bei der Differenzierung zwischen qualitativer und quantitativer Forschung allgemein (siehe oben § 1 D.I.) – tatsächlich „verbalisierende" bzw. „beziffernde" Forschungssynthesen gemeint. Um aber Verwechslungen von vornherein zu vermeiden, werde ich stattdessen von Forschungs*bericht* (1.) und Forschungs*auswertung* (2.) sprechen.

1. Forschungsbericht (narrative review)

Das klassische Instrument der Forschungssynthese ist der „erzählende" Forschungsbericht (*narrative review*). Er entspricht am ehesten dem, was in der Rechtswissenschaft als Kommentar bezeichnet wird: Ein Rezensent stellt relevante Erkenntnisse thematisch zusammen und diskutiert deren Verhältnis zueinander. Dabei muss er um Objektivität und Vollständigkeit bemüht sein, damit der Forschungsbericht als aussagekräftige Grundlage für weitere Forschung dienen kann.

Forschungsberichte sind meist ohne statistische Kenntnisse verständlich und auch für Laien leicht zu lesen. Sie „haben dort ihren Platz, wo es um die Kritik und den systematischen Vergleich von Theorien oder von inhaltlichen lnterpretationen empirischer Befunde geht."[205] Ihre große Stärke liegt darin, die historische Entwicklung eines Forschungsgebiets einschließlich seiner Querverbindungen zu anderen Gebieten geordnet darstellen zu können.[206] Forschungsberichte erzählen also Geschichten, das Attribut „narrativ" könnte daher kaum besser gewählt sein. Ebenso wie jeder Geschichtenerzähler – oder der Verfasser eines juristischen Kommentars – kann sich der Rezensent allerdings nicht davon befreien, durch Aufbau, Ablauf und Sprache seiner Darstellung subjektive Wertungen zu transportieren.[207] Überspitzt lässt sich deshalb fragen, warum eine empirische Disziplin, die ihren Gegenstand mittels quantitativer Methoden von subjektiven Urteilen befreien will, gerade im Rahmen der Forschungssynthese die-

---

[205] *Wagner/Weiß*, KZfSS Sonderheft 2004, 479, 481.

[206] *Ellis*, Effect Sizes 2010, 90: "useful for documenting the unfolding story of a particular research theme."

[207] *Hussy/Schreier/Echterhoff*, Forschungsmethoden 2010, 153: „Reviews verschiedener Wissenschaftler zum gleichen Thema können durchaus einen unterschiedlichen Forschungsstand vermitteln."

sen Anspruch plötzlich aufgeben und „unwissenschaftlich" vorgehen sollte.[208] Das hat vor allem drei Schwachstellen:

Erstens sind die von verschiedenen Primärstudien berichteten Zahlen in den seltensten Fällen direkt vergleichbar. Deshalb kann ein Forschungsbericht die Ergebnisse solcher Studien nur entweder anhand der Signifikanzwerte vergleichen – die ja nie das eigentliche Forschungsinteresse bilden (dazu unten B.IV.) – oder anhand der Attribute, mit denen die Studienautoren ihre Ergebnisse verbalisieren. Diese Attribute werden aber ebenso oft übertrieben wie Hypothesen nachträglich angepasst werden, um Ergebnisse beeindruckender darzustellen. Dabei ist selbst ohne solche Hintergedanken eine völlig neutrale und objektivierte Verbalisierung schlicht unmöglich. Lässt man Wissenschaftler beispielsweise zwanzig in der Fachliteratur veröffentlichte Versionen der Aussage „Therapie A ist effektiver als Therapie B." nach dem darin ausgedrückten Überzeugungsgrad sortieren, so mag die selbe Aussage, die dem einen als überzeugteste erscheint, dem anderen gerade am wenigsten überzeugt scheinen.[209] Welche Studie meint den größeren Effekt, wenn eine „deutliche" Anhaltspunkte für einen Effekt berichtet, die andere „starke"? Welcher Studie kommt mehr Gewicht zu, wenn die eine „deutliche" Anhaltspunkte für einen Anstieg, die andere aber „starke" Anhaltspunkte für einen Abfall berichtet? Hier hat der Rezensent praktisch nur drei Möglichkeiten: Er kann entweder nach Plausibilität entscheiden, also alle systematisch erhobene Empirie doch wieder an seiner eigenen subjektiven Willkür messen. Oder er ignoriert alle verbalen Umschreibungen, zählt stattdessen die signifikant positiven, signifikant negativen und die insignifikanten Ergebnisse und entscheidet nach einer Mehrheitsregel (*vote counting*); mathematisch lässt sich allerdings zeigen, dass dieses Vorgehen meist umso mehr Fehler erzeugt je mehr Studien der Rezensent berücksichtigt.[210] Drittens bleibt dem Rezensenten nur die Feststellung, der Forschungsstand sei unklar (*mixed evidence*) und weitere Forschung erforderlich – doch wieviel weitere Forschung? Wann reicht die Evidenz aus, um auch nur eine widersprechende Studie aufzuwiegen? Spätestens hier lässt sich eine willkürliche Grenzziehung nicht mehr vermeiden.

---

[208] *Bushman/Wells*, Pers Soc Psy Bull 2001, 1123: "It is somewhat ironic that the traditional review of scientific data has typically been conducted in an unscientific fashion. In the traditional narrative review, the reviewer uses 'mental algebra' to combine the findings from a collection of studies and describes the results verbally."; *Stanley*, J Econ Surv 2005, 309, 310: "non-experimental empirical economic research loses all claim to epistemic authority."

[209] So das Ergebnis des originellen Versuchs von *Dubben/Beck-Bornholdt*, Fehlinformation 2011, 189 ff.; vgl. auch *Abelson*, Principled Argument 1995, 54: "Some researchers […] call correlation coefficients of .35 'modest,' whereas others label them 'substantial.'"; aus Sicht des deutschen Rechts *Scholl*, NJW 1983, 319.

[210] *Hedges/Olkin*, Psy Bull 1980, 359.

Zweitens sind Rezensenten, die ein Feld hinreichend überblicken um es zu synthetisieren, fast immer auch selbst Primärforscher. Sie haben also ein handfestes Interesse daran, eigene Studien aufzuwerten und widersprechende Studien zu diskreditieren oder gar zu „übersehen".[211] Ebenso wie Replikationen selten veröffentlicht werden, wenn der Autor der Ausgangsstudie zum Gutachter bestellt wird (oben nach Fn. 192), setzen auch Forschungsberichte tendenziell persönliche Autorität an die Stelle überzeugender Methodik. Genau wie die Kreuzbegutachtung oft zu nachgerade zufälligen Ergebnissen führt (vgl. oben bei Fn. 204), weil sich immer genug Argumente für und gegen eine bestimmte Studie finden, können sich auch Forschungsberichte nicht davon frei machen, den Forschungsstand so darzustellen, wie der Rezensent ihn wahrnimmt.[212] Und warum sollte sich der Rezensent weniger von „hochsignifikanten" aber verschwindend geringen Effekten oder von „großen" Effekten bei noch größerer Streuung blenden lassen als Gutachter, die entsprechende Studien zur Veröffentlichung empfehlen?[213] Wer hier Rezensent ist, ist typischerweise dort Gutachter. Von einem Forschungsbericht darf also nicht erwartet werden, dass er fragwürdige Ergebnisse aussortieren oder einen besseren Qualitätsfilter abgeben kann als die Kreuzbegutachtung.

Drittens – und hier liegt der mit Abstand größte Mangel von Forschungsberichten – greifen Rezensenten nur auf veröffentlichte Studien zurück. Die Veröffentlichungslandschaft ist aber kein maßstabsgetreues Abbild der Forschungslandschaft, sondern vielfach systematisch verzerrt (oben VI.). Forschungsberichte vernachlässigen diese Verzerrungen, werden sie also eher perpetuieren als korrigieren.

2. Forschungsauswertung (meta analysis)

Die Alternative zum Forschungsbericht bildet die stärker formalisierte Forschungsauswertung, die auch als Metastudie bezeichnet wird (von griech.

---

[211] *Hussy/Schreier/Echterhoff*, Forschungsmethoden 2010, 153: „Unliebsame Studien, die das Gegenteil beweisen, werden [.] einfach nicht beachtet."; *Rosenthal/DiMatteo*, Ann Rev Psy 2001, 59, 62: "it may be all too tempting for authors of narrative reviews consciously or unconsciously to select and describe studies to support their own understanding"; *Ellis*, Effect Sizes 2010, 96: "Reading a narrative review one usually cannot tell whether it provides a full or partial survey of the literature. Were awkward findings conveniently ignored?"

[212] *Hussy/Schreier/Echterhoff*, Forschungsmethoden 2010, 153: „häufig zu beobachten ist, dass die Autoren solcher Artikel die Literatur oft [sic] so auswählen, dass ihre vorgefassten Schlussfolgerungen bestätigt werden."

[213] *Bushman/Wells*, Pers Soc Psy Bull 2001, 1123 demonstrierten etwa, dass die Ergebnisse von Studien mit besonders prägnantem Titel höher gewichtet werden als dieselben Ergebnisse unter einem weniger einprägsamen Titel.

μετά *meta*, „nach“, in der überkommenen Fehldeutung als „über, jenseits“),[214] weil sie nicht den für das Forschungsinteresse maßgeblichen Ausschnitt der Realität beobachtet, sondern die Forschung *über* diesen Realitätsausschnitt. Metastudien sind gewissermaßen „Studien über Studien“.[215]

a) Grundprinzip

Metastudien werten nicht beobachtete Rohdaten aus, sondern die Parameter aus früheren Auswertungen von beobachteten Rohdaten. Der oben (nach Fn. 101) beschriebene Abstraktionsschritt von den Rohdaten zum Parameter der beschreibenden Statistik vollzieht sich hier also erneut, von den Parametern verschiedener Rohdatensätze zu den Parametern der Metastudie. Das klingt umständlich – und ist es auch. Die ideale Metastudie würde nicht Parameter auswerten, sondern Rohdaten. Sie würde die Rohdaten aller früheren Studien zu einem Thema sammeln, sie zu einer riesigen Datenmenge vereinigen und diese Datenmenge auswerten.[216] Dass es solche idealen Metastudien nicht gibt, liegt einerseits daran, dass Forscher ihre Rohdaten selten veröffentlichen oder auf Nachfrage herausgeben (müssen), andererseits daran, dass veröffentlichte Parameter einfacher zu verwenden sind, weil sie halbwegs einheitlich berichtet werden, während jeder Forscher seine Rohdaten anders aufzeichnet (kodiert).

Das Vorgehen der idealen Metastudie verdeutlicht, welchen Zweck Metastudien im Kern haben: Sie sollen einen größeren Ausschnitt der interessierenden Grundgesamtheit beobachten und damit das Gewicht der schließenden Statistik im Erkenntnisprozess reduzieren. Paradoxerweise gelingt das gerade durch eine Anwendung der schließenden Statistik. Aus demselben Grund, aus dem Primärstudien Stichproben von mehr als einer Beobachtung ziehen, ziehen Metastudien Stichproben von mehr als einer Primärstudie: Je größer die Stichprobe, desto geringer die durchschnittliche Wertestreuung (Varianz) und das Gewicht einzelner Extremwerte (Ausreißer), und desto zuverlässiger erlaubt die schließende Statistik die Trennung systematischer von zufälligen Einflüssen.

---

[214] Vgl. Dudenredaktion, Fremdwörterbuch 2000, 865 („Metaphysik“): Als „das, was hinter der Physik (steht)“ (*tà metà tà physiká*) bezeichnete Andronikos von Rhodos „die philosophischen Schriften des Aristoteles, die in einer Ausgabe des 1. Jh.s v.Chr. hinter den naturwissenschaftlichen Schriften angeordnet waren“; weiter *Harper*, Etymology 2014, Stichworte “meta-” und “metaphysics”: Weil jene Schriften die Anfangsgründe des Seins *jenseits* der Physik behandelten, missverstanden spätere Interpreten die Vorsilbe so, wie es sich inzwischen eingebürgert hat.

[215] *Ellis*, Effect Sizes 2010, 94: “literally the statistical analysis of statistical analyses”.

[216] *Wagner/Weiß*, KZfSS Sonderheft 2004, 479 f. m.w.N. („gepoolte Auswertung“).

Dieses verkürzt dargestellte Grundprinzip sollte anschaulicher werden, wenn die Entstehung einer Metastudie Schritt für Schritt verfolgt wird. Weil Metastudien auch nur quantitativ-empirische Studien sind, gilt für sie dasselbe Sieben-Phasen-Schema, dessen siebte Phase hier gerade dargestellt wird.[217] „Eine Meta-Analyse umfasst sämtliche Elemente des sozialwissenschaftlichen Forschungsprozesses wie sie auch bei einer Primärforschung vollzogen werden",[218] deshalb unterliegen Metastudien auch weitgehend denselben Schwierigkeiten – insbesondere bei der statistischen Auswertung und bei der Veröffentlichung – wie Primärstudien. Ich verzichte daher auf deren Wiederholung und deute im nachfolgenden Ablaufschema nur zusätzliche, für Metastudien spezifische Aspekte an. Details finden sich wiederum in der einschlägigen Lehrbuchliteratur.[219]

b) Ablauf

Wie jede Forschungssynthese, beginnt die Metastudie mit einer *Sichtung der Literatur*. Diese ist allerdings wesentlich strenger formalisiert als etwa beim Forschungsbericht, weil die Metastudie zur Auswertung schließende Statistik einsetzt und deshalb sicherstellen muss, dass die ausgewertete Literatur sich von der tatsächlichen Forschungslandschaft nicht systematisch unterscheidet. Das bedeutet auch, dass die Metastudie sich nicht auf die veröffentlichte Literatur beschränken darf, um nicht deren Verzerrungen zum Opfer zu fallen. Deshalb werden in der Regel mindestens fünf Schritte unternommen, die jeweils exakt zu dokumentieren sind, um die Metastudie replizierbar zu machen:

Erstens wird in einschlägigen Datenbanken recherchiert wie unter I.2. beschrieben, allerdings ausgedehnt auch auf „graue Literatur wie

---

[217] In der Tat einschließlich der siebten Phase: Auch Metastudien bedürfen der Replikation (*Allen/Preiss*, J Soc Behav Pers 1993, 9, 11) und mittlerweile existieren zahlreiche Meta-Metastudien (z.B. *Lipsey/Wilson*, Am Psy'st 1993, 1181 über 156 Metastudien, beruhend auf 1,26 Mio. Teilnehmern; *Grissom*, J Cons Clin Psy 1996, 973 über 46 Metastudien; *Cafri/Kromrey/Brannick*, Multivar Behav Res 2010, 239 über 113 Metastudien; *Mitchell*, unten § 3 B.IV.3.; zur Methodik ausf. *Cooper/Koenka*, Am Psy'st 2012, 446) sowie Literaturberichte über Metastudien – Bsp. unten § 3 in Fn. 51 und Fn. 183, sowie *Richard/Bond/Stokes-Zoota*, Rev Gen Psy 2003, 331 mit den Ergebnissen von 322 sozialpsychologischen Metastudien, die 1898–1998 aus 25.000 Primärstudien mit 8 Mio. Teilnehmern entstanden waren.

[218] *Wagner/Weiß*, KZfSS Sonderheft 2004, 479, 483.

[219] Unübertroffen *Ellis*, Effect Sizes 2010, 89 ff. (Kap. 5 und 6); gründlich auch *Bortz/Döring*, Forschungsmethoden 2006, 671 ff. (Kap. 10) und *Sedlmeier/Renkewitz*, Forschungsmethoden 2008, 661 ff. (Kap. 22); sehr knapp *Finkelstein*, Basic Concepts 2009, 122 ff. und *Hussy/Schreier/Echterhoff*, Forschungsmethoden 2010, 153 ff.; ganz ausf. *Cooper/Hedges/Valentine*, Hdb Synthesis 2009; verbreitete Missverständnisse behandeln *Aguinis/Pierce u.a.*, Org Res Meth 2011, 306.

Institut[s]zeitschriften",[220] frei im Internet zugängliche Studien sowie etwaige Medien für Nullergebnisse (siehe oben bei Fn. 184).

Zweitens werden die Inhaltsverzeichnisse der letzten Jahrgänge aller für die Forschungsfrage besonders relevanten Zeitschriften manuell durchgesehen, um zu verhindern, dass relevante Texte übersehen wurden.

Drittens werden die im betreffenden Forschungsfeld besonders aktiven Forscher angeschrieben, um weitere (insbesondere unveröffentlichte) Studien ausfindig zu machen.

Viertens werden über einschlägige E-Mail-Verteiler, Internet- und andere Foren weitere unveröffentlichte Studien eingeworben, die dem Veröffentlichungsprozess zum Opfer gefallen sind.

Fünftens werden systematisch alle Querverweise verfolgt – sowohl durch eine vorwärtsgerichtete Verweissuche (oben I.2.) ausgehend von besonders relevanten Studien, als auch durch eine systematische Sichtung der Literaturverzeichnisse aller aufgefundenen Studien.

Am Ende dieser großangelegten Recherche steht eine Sammlung oft hunderter oder tausender Studien. Weil die nachfolgenden Schritte nicht minder aufwändig sind als der erste, empfiehlt es sich daher, eine gut durchdachte *Auswahl zu treffen*. Denn zum einen muss sichergestellt werden, dass die auszuwertenden Studien wirklich ähnliche Fragestellungen mit vergleichbaren Methoden bearbeiten, sonst weist die Metastudie ein „Äpfel-und-Birnen-Problem" auf (*apples and oranges*),[221] ist also nicht sinnvoll deutbar. Zum anderen ist es für eine Metastudie gar nicht erforderlich, durchweg alle bekannten Studien auszuwerten, sondern es genügt eine unverzerrte und hinreichend große Stichprobe.[222] In der Metastudie geht also Klasse vor Masse, wobei natürlich auch hier alle Auswahlkriterien ausdrücklich und nachvollziehbar niederzulegen sind.

Der danach verbleibende Grundbestand an Studien[223] ist sodann *detailliert zu kodieren* (katalogisieren). Dafür sind aus jeder Studie die relevanten Auswertungsparameter zu entnehmen oder nachträglich zu errechnen. Aufgrund der ganz unterschiedlichen Berichtsformate verschiedener Zeitschriften und der unterschiedlichen Aufbau- und Schreibstile verschiedener

---

[220] *Hussy/Schreier/Echterhoff*, Forschungsmethoden 2010, 155.

[221] *Hussy/Schreier/Echterhoff*, Forschungsmethoden 2010, 156; *Rosenthal/DiMatteo*, Ann Rev Psy 2001, 59, 68; *Wagner/Weiß*, KZfSS Sonderheft 2004, 479, 495 f.; *Ellis*, Effect Sizes 2010, 98; *Finkelstein*, Basic Concepts 2009, 123: "Two different studies rarely measure precisely the same parameter. […] One must then ask what one is testing or estimating when one combines a heterogeneous group of studies."

[222] *Ellis*, Effect Sizes 2010, 99 m.w.N.

[223] *Ellis*, Effect Sizes 2010, 99: "There could be anywhere from a few to several hundred studies in this group."

Autoren erfordert das nicht selten eine langwierige Suche im Text,[224] mitunter sogar eine Kontaktaufnahme mit den jeweiligen Autoren. Soweit sich bei der Durcharbeitung der Studien gravierende methodische Schwächen zeigen, müssen solche Studien gekennzeichnet oder entfernt werden, um das „Müll-rein-Müll-raus-Problem“ (*garbage in, garbage out*)[225] zu vermeiden, das die Aussagekraft der Metastudie beeinträchtigen würde.[226] Desweiteren werden nach der Kodierung oft diagnostische Instrumente verwendet, um verbleibende Verzerrungen der Stichprobe zu erkennen und bestenfalls zu korrigieren.[227]

Sind die relevanten Parameter jeder Studie katalogisiert, müssen nun die *Effektgrößen vereinheitlicht* werden. Die beiden bereits erwähnten Effektgrößenmaße (Mittelwertsunterschied und Korrelationskoeffizient, oben bei Fn. 115) sind nur die bekanntesten Vertreter zweier Familien – der Unterschiedseffekte (d-Familie) und der Zusammenhangseffekte (r-Familie) – zu denen noch viele weitere Maße gehören.[228] Oft wird gar keines berichtet, dann muss es erst aus anderen Parametern errechnet werden. Am Ende der Vereinheitlichung liegt für alle Studien jeweils ein Wert desselben Effektgrößenmaßes vor, also im Zweifel ein Maß für den Zusammenhang zweier Variablen, standardisiert an der Stichprobengröße und der Wertestreuung.

Dieses Maß sowie etwaige Kontrollvariablen gehen dann in die *statistische Auswertung* ein, die technisch vergleichsweise anspruchslos,[229] gleichwohl aber nicht frei von Herausforderungen ist.[230] Zuletzt werden die Ergebnisse der Metastudie gedeutet, was natürlich heißt, auf diejenigen

---

[224] *Ellis*, Effect Sizes 2010, 99: “Locating this information for a large set of studies may require hundreds of hours of careful reading.” und weiter auf 101: “coding is hard, mind-numbing work. It starts out being fun but often ends with the reviewer abandoning the project out of frustration or fatigue. Many of those who make it through the coding process never wish to repeat the experience.”

[225] *Rosenthal/DiMatteo*, Ann Rev Psy 2001, 59, 66 f.; *Wagner/Weiß*, KZfSS Sonderheft 2004, 479, 495 f.; *Hussy/Schreier/Echterhoff*, Forschungsmethoden 2010, 155 f.; *Ellis*, Effect Sizes 2010, 123.

[226] Das „gilt für jedes statistische Verfahren“ (*Bortz/Döring*, Forschungsmethoden 2006, 78), wird aber im Zusammenhang mit der Metastudie öfter betont als sonst.

[227] Zu diesen Instrumenten gehören etwa der Trichtergraph (*funnel plot*) und die kritische Gegenevidenzmasse (*fail-safe N*), näher *Ellis*, Effect Sizes 2010, 120 ff.; vgl. auch *Rosenthal/DiMatteo*, Ann Rev Psy 2001, 59, 66; *Wagner/Weiß*, KZfSS Sonderheft 2004, 479, 496 ff.; *Stanley*, J Econ Surv 2005, 309, 314 ff.; *Weiß/Wagner*, JBNST 2011, 661.

[228] *Rosenthal/DiMatteo*, Ann Rev Psy 2001, 59, 70 f.; ausf. *Ellis*, Effect Sizes 2010, 6 ff. mit tabellarischer Zusammenfassung über die Seiten 13–14.

[229] *Rosenthal/DiMatteo*, Ann Rev Psy 2001, 59, 69: “The level of quantitative skill and training required to do meta-analysis is very modest, [… with] few rather simple calculations needed to carry out a high-quality meta-analysis.”; *Ellis*, Effect Sizes 2010, 97: “the statistical analyses associated with meta-analysis are not difficult. If you can add, subtract, multiply and divide, you can combine effect sizes using a variety of approaches.”

Deutungen zu verzichten, die die Autoren jeder einzelnen Primärstudie ihren Daten angedacht hatten (vgl. noch unten bei Fn. 341). Diese haben für die Metastudie keinen Belang.

c) Vorteile

Aus dem Ablaufschema sollte bereits deutlich werden, dass Metastudien viele Nachteile von traditionellen Forschungsberichten vermeiden, und die meisten durch den Veröffentlichungsprozess verursachten Verzerrungen beheben.[231] Zusammengefasst lassen sich drei Hauptvorteile identifizieren:

> „Erstens bereichert die Metastudie den Vorgang der Forschungssynthese um ein hohes Maß an Disziplin. Viele Entscheidungen in diesem Vorgang sind subjektiv, aber anders als der Autor eines Forschungsberichts muss der Autor einer Forschungsauswertung diese Entscheidungen ausdrücklich treffen. […] Metastudien sind wie wissenschaftliche Buchprüfungen. Jeder Schritt muss aufgezeichnet, gerechtfertigt und für andere nachvollziehbar gemacht werden.
>
> Zweitens zwingen Metastudien, mit ihrem Schwerpunkt auf Datensammlung statt Datendeutung, den Autor dazu, sich sehr gut mit der Evidenz vertraut zu machen. Schlussfolgerungen aus Kurzzusammenfassungen zu entnehmen reicht nicht; der Autor muss die Methoden und Daten jeder einzelnen Studie beurteilen.
>
> Drittens, und das ist am wichtigsten, können Metastudien Fragen über die Beschaffenheit eines Effekts sogar dann klar beantworten, wenn sich die Befunde widersprechen.“[232]

Werden Metastudien verständig und kunstgerecht durchgeführt, stellen sie deshalb die beste bekannte Synthesemethode dar.[233] So haben mehrere empirische Studien (für die aber noch keine Synthese zu finden war) belegt, dass Forschungsauswertungen den Stand der Erkenntnis deutlich besser abbilden als Forschungsberichte.[234] Diese Erkenntnis wirkt sich unmittelbar auf die Stellung der Metastudie in der empirischen Erkenntnishierarchie aus, die ich unter B.VII. entwickeln werde.

---

[230] *Ellis*, Effect Sizes 2010, 117: “the real challenge is in identifying and dealing with multiple sources of bias.”; pointierte Übersicht bei *Aguinis/Pierce u.a.*, Org Res Meth 2011, 306.

[231] *Rosenthal/DiMatteo*, Ann Rev Psy 2001, 59, 61: “Meta-analysis allows researchers to arrive at conclusions that are more accurate and more credible than can be presented in any one primary study or in a nonquantitative, narrative review.”

[232] *Ellis*, Effect Sizes 2010, 96; ähnl. *Rosenthal/DiMatteo*, Ann Rev Psy 2001, 59, 63 ff., dort 64: “To extract the information needed to calculate effect sizes, a meta-analyst must become quite familiar with precisely what any given study actually found.”

[233] *Farley/Lehmann/Mann*, J Mktg Res 1998, 496, 501: “there is presently no better way to cumulate knowledge formally, as is evidenced by its use in life-or-death situations involving medical treatments.”

[234] *Cooper/Rosenthal*, Psy Bull 1980, 442; *Mann*, Sci 1994, 960; *Bushman/Wells*, Pers Soc Psy Bull 2001, 1123.

3. Theoriebildung

Eine letzte Form der Forschungssynthese ist die Theorie.[235] Als dritte Form der Forschungssynthese ans Ende der vorliegenden Darstellung gestellt zu werden, wird dem Stellenwert von Theorien in der empirischen Forschung eigentlich nicht gerecht. Theorien sind der Grund und das Ziel empirischer Forschung, die Quelle vieler Hypothesen und zugleich das eigentliche Erkenntnisinteresse hinter den meisten – erkundenden, beschreibenden oder schließenden – Studien.[236] Theorien helfen empirischen Forschern dabei, Hypothesen axiomatisch abzuleiten, also deduktiv zu arbeiten, und dadurch logisch unzulässige Induktionsschlüsse zu vermeiden:[237]

> „Die Gewinnung von Hypothesen aus Theorien (durch Deduktion) und die empirische Prüfung der Hypothesen (und damit der Theorie) ist der Normalfall in theoretisch-empirischen Wissenschaften. Diese Vorgehensweise entspricht auch dem deduktiv-empirischen Wissenschaftsmodell Poppers."[238]

Dieses philosophische Ideal kann die Theoriebildung freilich nur „in der Theorie" erfüllen. Praktisch arbeitet empirische Forschung allenfalls „hauptsächlich" deduktiv und passt Theorien nicht nur durch Falsifikation, sondern auch durch Exploration an die Realität an: „Im normalen Forschungsprozess spielt also immer beides eine Rolle: Induktion *und* Deduktion."[239] Zudem unterscheidet sich die Axiomatisierung und Formalisierung solcher Theorien (also auch ihr Stellenwert für die Aufstellung empirischer Hypothesen) zwischen verschiedenen Disziplinen ganz erheblich,[240] daher betrachte ich Theorien vorliegend nur in ihrer Synthesefunktion.[241]

In dieser Funktion ermöglichen Theorien den Schritt von der reduktionistischen Analyse zurück zur holistischen Schlussfolgerung und erfüllen damit eine wichtige Brückenfunktion.[242] Sie postulieren Zusammenhänge

---

[235] Statt aller *Walker/Willer* in: Webster/Sell, Experiments 2007, 52: "theory is a method of analysis *and* synthesis."

[236] Ausf. *Sedlmeier/Renkewitz*, Forschungsmethoden 2008, 39 ff.; *Bortz/Döring*, Forschungsmethoden 2006, 15 ff., 352 ff.; laut *Hussy/Schreier/Echterhoff*, Forschungsmethoden 2010, 3 ist die Theoriebildung sogar Definitionsmerkmal einer „empirischen Wissenschaft".

[237] *Hussy/Schreier/Echterhoff*, Forschungsmethoden 2010, 8; zum Induktionsproblem kurz *Eidenmüller*, JZ 1999, 53, 54; ausf. *Ulen*, U Ill L Rev 2002, 875, 882 ff.

[238] *Diekmann*, Sozialforschung 2012, 190; Werkauswahl in *Popper*, Problemlösen 2001.

[239] *Sedlmeier/Renkewitz*, Forschungsmethoden 2008, 25.

[240] *Engel* in: Engel u.a., Recht und Verhalten 2007, 363, 376 ff.; *Camerer*, PNAS 1999, 10575: "to an economist, a theory is a body of mathematical tools and theorems. To a psychologist, a theory is a verbal construct or theme that organizes experimental regularity."

[241] Ausf. zur juristischen Rezeption sozialwissenschaftlicher Theorie *Engel* in: Engel, Methodische Zugänge 1998, 11.

[242] So auch *Mook*, Am Psy'st 1983, 379, 384 f.; *Walker/Willer* in: Webster/Sell, Experiments 2007, 51 f.; anders akzentuiert *Ietswaart* in: Plett/Ziegert, Empirische Rechtsforschung 1984, 210, 211: Brücke zwischen "mere information and meaningful knowledge."

zwischen unterschiedlichen empirischen Erkenntnissen und bestimmen die Geltungsbedingungen und Grenzen solcher Erkenntnisse. Auf lange Sicht sollten Theorien also die empirischen Erkenntnisse über einen größeren Realitätsausschnitt zusammenfassen und systematisieren. Dadurch erleichtern sie es dem Rezipienten, einen Überblick zu gewinnen, tragen aber zugleich ein bestimmtes Vorverständnis an empirische Untersuchungen heran,[243] das den Raum der möglichen Erkenntnisse beschränkt:

> „Theorien sind Modelle des Phänomens, welches erklärt werden soll: eine Beschreibung des Phänomens, welches von allen nicht essentiellen Einzelheiten befreit ist – ungefähr so, wie ein Stadtplan eine Beschreibung einer Stadt ist, die einen führen soll. Je nach Zweck des Plans werden nur die Hauptstraßen eingezeichnet, manchmal andere Straßen, aber jedenfalls nicht Fußgängerüberwege und Ampeln. Andere Stadtpläne, etwa für eine Baufirma, die Kanalisationen baut, müssen dagegen Erhöhungen, Elektrizitätskabel etc. kennzeichnen."[244]

Die Entscheidung, welche „Einzelheiten" für eine Theorie „essentiell" sind, ist freilich eine wertende. Der Rezipient muss implizite normative Annahmen einer Theorie deshalb kritisch hinterfragen und unter Umständen durch eigene Wertungen ersetzen (dazu schon oben bei Fn. 167). Soweit dies gelingt, kann aber auch die Theorie als eine Art der *verstehenden* Forschungssynthese gewinnbringend rezipiert werden.

B. Grundsätze der empirischen Rezeption

Aus dem eben dargestellten Forschungsablauf lassen sich sieben pragmatische Grundsätze (Faustregeln) ableiten für die juristische Rezeption empirischer Erkenntnis. Davon sind der erste und letzte eher grundsätzlicher Natur, während die mittleren fünf dabei helfen sollen, die Ergebnisse einzelner Primärstudien zu würdigen. Die nachfolgenden Ausführungen vertiefen also das bereits Dargelegte unter dem stärker anwendungsbezogenen Blickwinkel des Rezipienten.

---

[243] Ausdr. *Atteslander*, Methoden 2010, 21: „Theorien [sind] im Grunde Entscheidungen über die Bedeutung und Bedingungen von erfassbaren Erscheinungen der sozialen Wirklichkeit."

[244] *van Aaken*, Rational Choice 2003, 37; ebenso *Lawless/Robbennolt/Ulen*, Methods 2010, 9; *Diekmann*, Sozialforschung 2012, 145: „Die Landkarte soll die Wirklichkeit nicht fotografisch abbilden, sondern die wesentlichen Merkmale und Zusammenhänge hervorheben. Je nach Zielsetzung wird man […] einen mehr oder minder feinen Maßstab wählen."

I. Alle empirische Forschung ist implizit normativ.

Der für die Rezeption empirischer Forschung vielleicht wichtigste Grundsatz betrifft die Anschlussfähigkeit der Rechtswissenschaft an empirische Disziplinen. Er beruht auf der Erkenntnis:

„‚Reine' Empirie, die voraussetzungslos die Wirklichkeit auf den Begriff bringt, ist aus erkenntnistheoretischen Gründen unmöglich. Alle Empirie ist durch Sprache, Beobachtungsgewohnheiten, wissenschaftliche Theorien und Hypothesen bewußt oder unbewußt geleitet. Wirklichkeit begegnet uns – wissenschaftlich feststellbar – immer nur unter solchen leitenden Gesichtspunkten."[245]

Empirische Erkenntnisse bilden die Wahrheit also nicht objektiv ab, sondern transportieren immer Wertungen.[246] Diese Wertungen zu erkennen und unter einem juristischen Blickwinkel zu würdigen – „sowohl mit Respektlosigkeit als auch mit angemessener wissenschaftlicher Ernsthaftigkeit"[247] –, bildet die wohl fruchtbarste Rezeptionsstrategie. Schließlich sind Juristen dank ihrer intimen Vertrautheit mit den Institutionen des praktischen Lebens in einer hervorragenden Position, um die Problemstrukturen herauszuarbeiten, denen die empirische Forschungsplanung Rechnung tragen muss und von denen sie nicht abstrahieren kann, ohne die Problemdefinition nennenswert zu ändern und die Gültigkeit der so gewonnenen Erkenntnis in Frage zu stellen.[248]

Dabei stellt sich zunächst das *Relevanzproblem*:

„Die Menge möglicher Forschungsprobleme ist unendlich, Zeit und materielle Ressourcen […] begrenzt. In welche Richtung das Interesse der Forschung gelenkt wird, […] ist im höchsten Maße ein Wertproblem."[249]

---

[245] *Starck*, JZ 1972, 609, 614 (These 3); *Augsberg*, Staat 2012, 117, 125 zur „Empirie an sich"; ebenso ist nach *Arndt*, Empirie 2008, 45 „schon die Entscheidung für die Betrachtung eines Wirklichkeitsausschnittes subjektiv: Auswahl, Beschreibung und Interpretation von tatsächlichen Vorgängen […] hängen [.] von vielfältigen individuellen Faktoren ab";

[246] *Langenbucher*, ZGR 2012, 314, 315 unter dem Schlagwort „Vorverständnisprägung" (freilich ihrerseits durch ein wissenschaftsphilosophisches Vorverständnis geprägt); zu eng daher *Petersen*, Staat 2010, 435, 447: „enthalten auch *viele* empirische Studien implizite normative Wertungen" (Hervorhebung nur hier); vgl. *Augsberg*, Staat 2012, 117, 118.

[247] *Atteslander*, Methoden 2010, 23.

[248] *Ho/Rubin*, Ann Rev L Soc Sci 2011, 17, 34: "research should empower the broader legal academic community—precisely the community with the comparative advantage—to assess the credibility of the inference."; *Schneider/Teitelbaum*, Utah L Rev 2006, 53, 66: "Legal academics, because of their legal training, can perceive and grasp some critical things about legal institutions better than lay scholars."; *Neumann/Krieger*, Clin L Rev 2003, 349, 390: "researchers with only a social science background (and no legal training and experience) often produce scholarship that […] suffers from naiveté about law and its institutions."

[249] *Diekmann*, Sozialforschung 2012, 80 (zu den anderen drei Aspekten des „Werturteilsproblems" vgl. 76 ff.).

Was Juristen empirisch untersuchungswürdig erscheint, muss sich keineswegs mit dem decken, was empirische Forscher tatsächlich untersuchen (vgl. schon § 1 bei und in Fn. 208). Insbesondere ist empirische Forschung oft weniger einer in der Lebenswirklichkeit vorgefundenen Problemstruktur verpflichtet als vielmehr der Verwirklichung ihres Methodenpotentials. Nach dem Kaplan-Maslowschen Hammerprinzip[250] wird der Untersuchungsgegenstand oft so definiert und abgegrenzt, wie er sich mit dem empirischen Instrumentarium am besten untersuchen lässt.[251] Das trägt zwar sicher dazu bei, dass das empirische Endprodukt hohen methodischen Standards genügt. Umso weiter aber ist es dem direkten Zugriff des Rezipienten entrückt, weil zusätzlich zur eigentlichen Problemstruktur zahlreiche Annahmen und implizit normative Entscheidungen erforderlich wurden.

Solche implizit normativen Entscheidungen finden sich an vielen Stellen des empirischen Verfahrens, vom gewählten Gültigkeitskriterium (dazu unten § 3 A.I.), über die Operationalisierung von Variablen,[252] bis hin zur Definition relevanter Effektgrößen[253] und den Methoden der statistischen Auswertung (siehe oben § 2 A.IV.2.a) – „Werturteile und Interessen können sich bei diesen Entscheidungen bemerkbar machen.“[254] Um sie kritisch hinterfragen zu können, muss „der juristische Rezipient […] die normativen Implikationen herausarbeiten, die in dem (oft ebenfalls unausgesprochenen oder bloß angedeuteten) analytischen Apparat versteckt sind.“[255] Er

[250] *Kaplan*, Inquiry 1964, 28: “Give a small boy a hammer, and he will find that everything he encounters needs pounding.” (ähnl. ebd., 11 mit der Geschichte vom Betrunkenen, der seinen Schlüssel im Dunklen verliert, aber unter der Straßenlaterne sucht, weil es dort heller sei.); *Maslow*, Science 1966, 15: “I suppose it is tempting, if the only tool you have is a hammer, to treat everything as if it were a nail.”

[251] Krit. schon *Nußbaum*, AcP 1955, 453, 478: „Die Statistik-Gläubigkeit […] neigt dazu, die Aufmerksamkeit auf Punkte zu lenken, die statistisch erfaßbar sind oder als erfaßbar gelten; und daher andere, häufig wichtigere Probleme zu vernachlässigen“; *Schneider/Teitelbaum*, Utah L Rev 2006, 53, 66: “social scientists’ research is usually driven by the theoretical concerns and traditional topics of their disciplines, not what lawmakers need to know.”

[252] *Petersen*, Staat 2010, 435, 451 ff.; *Beller*, Forschen lernen 2008, 17: „Operationalisierung ist in letzter Konsequenz immer eine normative Entscheidung, mit der gleichzeitig der Raum möglicher Ergebnisse definiert wird.“

[253] *Beller*, Forschen lernen 2008, 110: „Die Bestimmung der (Mindest-)Größe eines praktisch bedeutsamen Effekts erfordert – wie die Festlegung des α- und β-Niveaus – eine *inhaltliche* Auseinandersetzung mit der Fragestellung und ist keine statistische Frage.“; sie könne „nur durch eine ‚Güterabwägung‘ aufgrund inhaltlicher Überlegungen geklärt werden. Die verlangt letztendlich eine normative Entscheidung“ (104).

[254] *Diekmann*, Sozialforschung 2012, 83; *Mastronardi*, Juristisches Denken 2003, Rn. 67: „Alle Sozialwissenschaft ist […] ein (oft verborgenes) Stück weit normativ.“; noch kritischer *Simmons/Nelson/Simonsohn*, Psy Sci 2011, 1359 mit dem Untertitel “Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant”.

[255] *Engel* in: Engel u.a., Recht und Verhalten 2007, 363, 388; *Petersen*, Staat 2010, 435, 447.

bedarf also eines hinreichenden Verständnisses der empirischen Forschungslogik und Methoden, wie es die vorliegende Darstellung zu vermitteln sucht. Skepsis und Argwohn allein genügen ebenso wenig wie blindes Vertrauen auf Statistik und Wissenschaftsstandards. Auf dem Basar der empirischen Forschung ist kritische Urteilskraft die Währung.[256]

II. Sorgfältige Planung geht vor statistischer Raffinesse.

An verschiedenen Stellen der vorangegangenen Darstellung wurde die Bedeutung der Studienplanung betont. Beispielsweise lassen sich Ursachen überhaupt nur durch einen bestimmten Studienentwurf, niemals aber durch die Datenauswertung isolieren (oben bei Fn. 53) und hängt die Bedeutung signifikanter Ergebnisse von der vorab ermittelten Teststärke ab (oben bei Fn. 82). Daraus erklärt sich der zweite Rezeptionsgrundsatz, der andernorts noch kürzer gefasst wurde: „Studienplanung übertrumpft Auswertungsmethoden".[257]

Die Datenauswertung mittels der Statistik kann eine reflektierte Deutung der Daten nur unterstützen und vorbereiten, aber nie ersetzen.[258] Oft genug wurde darauf hingewiesen, dass aus „schlechten" (fehlerhaft erhobenen) Daten auch durch die beste statistische Auswertung niemals „gute" (aussagekräftige replizierbare) Ergebnisse zu gewinnen sind.[259] Statistiksoftware macht noch keinen Statistiker, genau wie das Skalpell noch kei-

[256] *Gorard*, Brit J Sociol Edu 2006, 67, 77: "research is not typically taught as an exercise in judgement […] judgement seems 'subjective' whereas computation is ostensibly 'objective'."; *Kaplan*, Inquiry 1964, 29: "statistical formulas are but instruments after all; it is not they that produce scientific results but the investigator who uses them scientifically."

[257] *Ho/Rubin*, Ann Rev L Soc Sci 2011, 17, 22: "research design trumps methods of analysis."; ähnl. *Angrist/Pischke*, J Econ Persp 2010, 3 zur "Credibility Revolution in Empirical Economics"; prägnant auch *Guttman*, Appl Stoch M D A 1985, 3: das „Gerüst" der Statistik müsse sich dem „Gebäude" der Wissenschaft anpassen; vgl. schon oben bei Fn. 129.

[258] *Nickerson*, Psy Meth 2000, 241, 290 f.: "There are no statistical procedures that can safely be used without thought […] Statistical methods should facilitate good thinking, and only to the degree that they do so are they being used well."; *Neumann/Krieger*, Clin L Rev 2003, 349, 391: "it leads in false directions when a researcher substitutes statistical gymnastics for an astute understanding of the human and social elements of a problem."; nichts anderes meinen *Angrist/Pischke*, Econometrics 2009, 8: "Although inference issues are rarely very exciting, and often quite technical, the ultimate success of even a well-conceived and conceptually exciting project turns on the details of statistical inference."

[259] *Aiken*, Edu Psy Meas 1994, 848, 850: "Dressing up meaningless data in the clothing of complex statistical methodology cannot transform a body of evidence into something of value."; *Gorard*, Brit J Sociol Edu 2006, 67, 76: "Complex statistical methods cannot be used *post hoc* to overcome design problems or deficiencies in datasets."; *Lawless/Robbennolt/Ulen*, Methods 2010, 26: "without appropriate data, the available statistical techniques are meaningless."; in der Physik können vor allem Himbeerkuchen die Datenerhebung gefährden, so *Heinicke*, Velocimetry 2013, 56.

nen Chirurgen macht[260] – und „das Resultat unüberlegter und mangelhaft geplanter empirischer ‚Forschung' [ist] nicht selten ein kaum noch genießbarer Datensalat und frustrierte Forscher oder Forscherinnen."[261] Dennoch „hat sich die einstmalige Zahlenfeindlichkeit in eine fast schon übertriebene, zumindest äußerliche Zahlenverehrung […] umgekehrt."[262] Die viel wichtigeren Fragen der Studienplanung treten demgegenüber oft in den Hintergrund.

Illustrieren lassen sich diese Überlegungen an der empirischen Corporate-Governance-Forschung, die sehr oft die beeindruckende Größe ihrer Datensätze betont und komplizierte statistische Verfahren einsetzt, aber nicht reflektiert, welche Konsequenzen die zur Gewinnung dieser Datensätze angewandten Methoden und welche Grenzen die statistische Auswertung haben.[263] Woraus bilden die untersuchten Unternehmen eine Stichprobe? Aus allen Unternehmen? Aus allen Aktiengesellschaften? Aus allen börsennotierten Aktiengesellschaften? Aus allen indexnotierten Aktiengesellschaften? Oder aus allen publikationsfreudigen indexnotierten Aktiengesellschaften? Für welche Klasse von Unternehmen, welchen Zeitraum und welche Fragestellungen beanspruchen die gewonnenen Erkenntnisse Geltung, und auf Grundlage welcher Überlegungen zur Übertragbarkeit? Wenn diese Fragen nicht reflektiert werden und Studien stattdessen mit der größtmöglichen Zahl leicht verfügbarer (Kapitalmarkt-) Daten durchgeführt werden, haben ihre Erkenntnisse oft nur eine sehr beschränkte Aussagekraft.

Deshalb gilt auch bei der Auswahl der Stichprobe das Prinzip „Klasse statt Masse",[264] denn die Stichprobengröße reduziert zwar den Einfluss von Zufallsschwankungen, erhöht aber nicht zwangsläufig die Übertragbarkeit

---

[260] *Good/Hardin*, Errors 2009, S. xi: "statistical software will no more make one a statistician than a scalpel will turn one into a neurosurgeon. Allowing these tools to do our thinking is a sure recipe for disaster."; *Rosenthal/DiMatteo*, Ann Rev Psy 2001, 59, 68: "'high-tech statistication' […] lend[s] an impressive air of sophistication but may be massively inappropriate."

[261] *Diekmann*, Sozialforschung 2012, 187; *Epstein/Martin* in: Cane/Kritzer, Hdb Empirical 2010, 901, 905: "even the best statistician cannot make lemonade without lemons."

[262] *Krämer*, KZfSS Sonderheft 2004, 51 m.w.N., und auf 59: „Falsch verstandene klassische Inferenzstatistik ist schlechter als gar keine Inferenzstatistik."

[263] Eindringlich *Devers/Cannella u.a.*, J Mgmt 2007, 1016, 1041: "the perpetuation of a multitude of analytic tools that are not fully understood or justified, and their corresponding disparate results, presents a slippery slope that we strongly caution against."; vgl. auch § 3 B.II.2.b) zum Selektionsfehler in Beobachtungsstudien.

[264] *van Belle*, Rules of Thumb 2008, 55: "The adage, 'there is strength in numbers' may be true in politics but less likely to be so in science. There is more strength in fewer but well-chosen numbers."; gutes Bsp. bei *Sedlmeier/Renkewitz*, Forschungsmethoden 2008, 120.

der Daten – das kann nur der Auswahlprozess.[265] Dementsprechend viel Gewicht muss bei der Durchführung empirischer Studien auf die Planungsphase gelegt werden, und genauso viel Aufmerksamkeit sollte der Rezipient darauf verwenden, die Annahmen und Entscheidungen der Studienplanung zu hinterfragen, denn jede „Analysestrategie" ist

> „mit eigenen Annahmen verbunden, die sich nicht immer direkt überprüfen lassen. Wichtig ist vielmehr eine adäquate Auseinandersetzung mit diesen Annahmen und dem entscheidenden Selektionsprozess."[266]

III. Ohne Hypothese kein Ergebnis.

Der Vorrang der Planung vor der Auswertung führt unmittelbar zum dritten Rezeptionsgrundsatz, der an verschiedenen Stellen schon vorweggenommen wurde (z.B. oben bei Fn. 85 und Fn. 161). Er bezieht sich auf die schließende Statistik, die aber in der internationalen empirischen Forschung eine überragende Bedeutung hat und auch in Deutschland zunehmend relevant wird.

Der Glaube an die statistische Allmacht (vgl. oben bei Fn. 251 und Fn. 262 sowie unten bei Fn. 288) findet reichlich Bekräftigung durch die Leistungsfähigkeit moderner Computer. Sie nehmen jede denkbare Art von Daten, führen beliebige Berechnungen beliebig oft durch und geben (fast) immer ein Ergebnis aus. Die Verlockung ist deshalb groß, beliebige Daten zu sammeln und statistisch zu durchforsten, um Regelmäßigkeiten zu entdecken (*data mining*). Die Statistik-Software benötigt keine Hypothese, um einen Mittelwertsunterschied und Signifikanzwert auszurechnen.[267] Doch gerade *weil* das so ist, kommt empirische Stichprobenforschung nicht ohne vorab definierte Hypothesen aus. Solche Hypothesen trennen erst die Spreu vom Weizen, erkundende von schließender Statistik, anekdotische Evidenz von empirischer Erkenntnis.

Ein berühmtes Beispiel sind die Kalendereffekte am Aktienmarkt:[268] Man könnte unzählige mögliche Zusammenhänge untersuchen zwischen Tageszeit, Datum, Jahreszeit, Jahreszahl, usw. einerseits und dem Börsenpreis allgemein, in bestimmten Ländern, Branchen, Anfangsbuchstaben, usw. andererseits. Die Anzahl der denkbaren Kombinationen geht gegen

---

[265] *van Belle*, Rules of Thumb 2008, 56: "Sample size determines precision not accuracy. The selection process determines accuracy (or validity)."

[266] *Legewie*, KZfSS 2012, 123, 140.

[267] *Mintken*, VR 1992, 252, 253: „Es ist wenig sinnvoll, Analysen nur deswegen durchzuführen, weil das benutzte Rechenprogramm sie erlaubt. Das führt zwar in der Regel zu einer beachtlichen Menge an Kennzahlen, Tabellen und Grafiken, jedoch ist der inhaltliche Ertrag zumeist gering."; vgl. schon eben bei und in Fn. 260.

[268] Gemeint ist „die Erheblichkeit bestimmter zeitbedingter Muster im Kursverlauf, wie den Montagseffekt und den Januareffekt", *Benicke*, ZGR 2004, 760, 766.

unendlich, und für jede zwanzigste ergibt die herrschende statistische Methodik ein signifikantes Ergebnis – selbst dann, wenn kein einziger Zusammenhang zwischen Kalender und Börsenpreis besteht. Da überrascht es nicht, dass tatsächlich Kalendereffekte beobachtet werden, und darunter auch solche, die man plausiblerweise hätte erwarten können – etwa den berüchtigten Montagseffekt.[269] Man mag gute und plausible Erklärungen dafür finden, warum der Börsenpreis montags deutlich niedriger sein sollte als freitags. Ebenso gute Erklärungen fände man aber auch für einen höheren Preis, oder einen Dienstags-, Mittwochs- oder sonst einen Effekt (vgl. oben Fn. 27). Untersucht man dagegen alle möglichen Kalendereffekte, verschwinden plötzlich die überzufälligen Regelmäßigkeiten; der vermeintliche Montagseffekt ist nur Teil des statistischen Grundrauschens.[270] Ohne (oder mit nachträglich gebildeten) Hypothesen läuft jede empirische Erhebung Gefahr, Grundrauschen für bedeutungsvoll zu halten, weil es plausibel erscheint. Plausibilität ist aber gerade kein Kriterium für die Datenauswertung (vgl. oben bei Fn. 18), sonst bedürfte es der Datenauswertung gar nicht.

Leider fällt es sehr schwer, behauptete Hypothesen darauf zu überprüfen, ob sie bereits vor der Erhebung existierten (vgl. schon oben bei Fn. 163). Wenn sich nicht klären lässt, ob die Hypothese vorab feststand, lässt sich allenfalls als Faustregel formulieren: Je weniger naheliegend eine Hypothese erscheint, desto gründlicher sollte sie aus vorherigen Erkenntnissen und Theorien hergeleitet sein.[271] Das entspricht auch juristischen Denkgewohnheiten.[272] Rezipienten empirischer Forschung sollten deshalb den Prozess (nicht das Ergebnis!) der Hypothesenbildung immer kritisch hinterfragen. Wirklich abhelfen kann „verspäteten" Hypothesen letztlich nur die Replikation, denn Replikationsstudien gehen per Definition immer von einer Hypothese aus – nämlich dem Ergebnis der Ursprungsstudie. Lässt sich dann der erwartete Effekt erneut belegen, spricht einiges für seine Existenz. Leider werden aufgrund der oben dargelegten systematischen Verzerrung des Veröffentlichungsverfahrens (§ 2 A.VI.3.) praktisch nie reine Replikationen veröffentlicht.[273]

---

[269] *Triebe*, NZZ 9. August 2010: „Der Montagseffekt besagt, dass sich die Aktienrendite vom Schlusskurs am Freitag bis zum Schlusskurs am folgenden Montag durchschnittlich schlechter entwickelt als im Rest der Woche."

[270] *Sullivan/Timmermann/White*, J Econometrics 2001, 249 m.w.N.

[271] Um Missverständnisse zu vermeiden: *Wenn* eine Hypothese vorab feststand, ist ganz egal, woher sie kam. Die vorliegende Faustregel begegnet nur der Unsicherheit, *ob* eine Hypothese vorab feststand.

[272] *Engel*, JITE 2010, 199, 201: "the more one deviates from what most other lawyers think is the appropriate way of solving the case, the stronger the arguments one needs."

[273] *Nickerson*, Psy Meth 2000, 241, 283: "Experiments that are literal replications of previously published experiments are very seldom published—I do not believe I have ever seen

IV. Ergebnis ist nicht die Signifikanz, sondern die Effektgröße.

Die vorige Rezeptionsregel besagte unmittelbar nur, dass hypothesenfrei durchgeführte Signifikanztests nicht sinnvoll deutbar sind. Damit sollte aber nicht gesagt sein, dass der sinnvoll deutbare Signifikanztest auch schon das empirische „Ergebnis“ darstellt. Viele empirische Forscher, Schriftleiter und Gutachter hingegen verengen den Blick auf den Signifikanztest. Signifikanz kommt von lat. *significantia*, Bedeutung,[274] also ist Signifikanz ein Maß für die Bedeutung einer Erkenntnis – so das weit verbreitete Verständnis.[275] Ein Missverständnis, leider.[276]

Obwohl es sich bei der Signifikanz „nie um eine Aussage über die Effektgröße, die Relevanz des Ergebnisses oder des Unterschiedes“ handelt,[277] werden substantiell bedeutsame Erkenntnisse oft als „signifikant“ bezeichnet. Das ist Folge eines Bedeutungswandels, den dieses Wort im zwanzigsten Jahrhundert erfahren hat, nachdem es in der Statistik bereits als Fachbegriff gebräuchlich war.[278] Die statistische Bedeutung des Wortes lässt sich leicht erfassen, wenn man erkennt, dass die lateinische Grundform von *significantia* zusammengesetzt ist aus *signum*, Zeichen, und *facere*, machen:[279] Signifikanz ist ein Anzeichen dafür, dass ein Stichprobeneffekt auch in der Grundgesamtheit existiert, nicht mehr und nicht weniger.

Das wird daran deutlich, dass die meisten Nullhypothesen tatsächlich als Punkthypothesen der Art „Es gibt keinen Unterschied, der Effekt ist exakt null.“ (*nil hypothesis*) formuliert werden, dass aber zugleich zwei aus

---

one.“; anders akzentuiert *Dubben/Beck-Bornholdt*, Fehlinformation 2011, 104: „Fremde Experimente und Studien zu wiederholen und damit zu überprüfen, wird [.] von Wissenschaftlern als langweilig empfunden und ist mit dem Verdacht verbunden, man habe keine eigenen erforschenswerten Ideen“.

[274] *Dudenredaktion*, Fremdwörterbuch 2000, 1229 („Signifikanz“); *Köbler*, Lateinisches Wb. 2009; *Harper*, Etymology 2014, Stichwort “significance”.

[275] *Vickers*, Understand Statistics 2010, 148: “perhaps the most typical approach to statistics”; pointiert *Krämer*, Denkste! 2011, 187: „Für viele ist ‚Signifikanz‘ eine Art Adelstitel: wissenschaftlich untermauert, empirisch unangreifbar, jenseits allen Zweifels abgesichert, die TÜV-Plakette der modernen Datenhändler. Aber diese Assoziationskette signifikant = wichtig = richtig ist falsch.“

[276] Ausf. *Bortz/Döring*, Forschungsmethoden 2006, 600; *Sedlmeier/Renkewitz*, Forschungsmethoden 2008, 782 f.; *Ziliak/McCloskey*, Statistical Significance 2008, passim; *Seth/Carlson u.a.* in: Bergh/Ketchen, Research Methodology 2009, 3; *Ellis*, Effect Sizes 2010, 3; spezifisch juristisch *Lempert*, L Soc Inq 2009, 225; *Zeisel/Kaye*, Figures 1997 nennen das die “grossness fallacy”.

[277] *Petersen/Goerg* in: Towfigh/Petersen, Methoden 2010, 201, 226.

[278] *Salsburg*, Statistics 2001, 98: “The word was used in its late-nineteenth-century English meaning, which is simply that the computation signified or showed something.”

[279] *Dudenredaktion*, Fremdwörterbuch 2000, 1229 („signifizieren“); *Köbler*, Lateinisches Wb. 2009; *Harper*, Etymology 2014, Stichwort “signify”.

realen Daten gezogene Mittelwerte niemals exakt identisch sein werden.[280] Deshalb „ist jede Nullhypothese letztlich chancenlos, wenn man die eingesetzten Stichproben genügend groß macht, d. h., Signifikanz ist letztlich eine Frage des Stichprobenumfanges.“[281] Umgekehrt formuliert: *Insignifikanz* bedeutet im Prinzip nur, dass die Stichprobe nicht hinreichend sicher erkennen lässt, ob der betrachtete Effekt ein positives oder negatives Vorzeichen trägt,[282] und diese Unsicherheit kann ebenso darauf beruhen, dass die Stichprobe zu klein ist um das Vorzeichen zu bestimmen, wie darauf, dass der Effekt kein Vorzeichen *hat*, weil er nicht existiert (null ist). Die Rezeption von Signifikanzwerten muss deshalb die Stichprobengröße im Auge behalten: Kleine Stichproben können selbst Mammuteffekte als insignifikant erscheinen lassen, große Stichproben dagegen können selbst minimale Effekte bis zur Signifikanz aufblasen.[283]

Diese Erkenntnisse verdeutlichen zweierlei: Zum einen, dass Hypothesen so präzise formuliert sein sollten wie möglich, weil andernfalls die Widerlegung der Nullhypothese geradezu unausweichlich ist und keine neue Erkenntnis bringt. Zum anderen, dass Signifikanz aufgrund ihrer Stichprobenabhängigkeit nicht das (einzige) Kriterium sein kann, um einen Effekt in der Grundgesamtheit zu beurteilen.[284] Wir erinnern uns: Könnten wir die gesamte Grundgesamtheit beobachten, könnten wir gleich die Effektgröße berechnen und bräuchten keine schließende Statistik. (In Vollerhebungen versteht sich die vorliegende Rezeptionsregel deshalb von selbst.) Schließende Statistik ist nur der notwendige Zwischenschritt, weil wir die Grundgesamtheit nicht vollständig beobachten können.[285] Deshalb ändert sich aber nicht gleich das Erkenntnisinteresse: Auch in einer Stichprobenerhebung ist die Stärke eines Effekts das eigentlich relevante Ergebnis. Während die Signifikanz je weniger aussagt desto größer die Stichprobe

---

[280] *Bortz/Döring*, Forschungsmethoden 2006, 28, 600 sprechen von einer „reinen Fiktion“.

[281] *Bortz/Döring*, Forschungsmethoden 2006, 28, 600 ff.; „Mit anderen Worten, bei großen Stichproben sind selbst kleinste Abweichungen von der Ausgangshypothese schon ‚signifikant‘, bei kleinen Stichproben dagegen kann man selbst grobe Verletzungen der Ausgangshypothese noch dem Zufall in die Schuhe schieben.“ (*Krämer*, Denkste! 2011, 190); vgl. auch *Gorard*, Brit J Sociol Edu 2006, 67, 71.

[282] So *Shadish/Cook/Campbell*, Causal Inference 2002, 44, 52 m.w.N.; *Abelson*, Principled Argument 1995, 74: “the so-called acceptance of the null hypothesis merely signifies a reluctance to bet on the direction of the true mean difference.”; gleichbedeutend mit der Erläuterung oben bei Fn. 159.

[283] Zu einem Beispiel, bei dem man durchaus zweifeln kann, ob die Signifikanz nicht nur Resultat einer sehr großen Stichprobe ist, unten § 4 Fn. 589.

[284] *Rosenthal/DiMatteo*, Ann Rev Psy 2001, 59, 63 (“Keeping Statistical Significance In Perspective”); *Beck-Bornholdt/Dubben*, Nature 1996, 730 weisen darauf hin, dass sich andernfalls mittels der Signifikanz sogar im simplen Dreischritt beweisen ließe, dass der Papst ein Außerirdischer ist.

[285] So auch *Sedlmeier/Renkewitz*, Forschungsmethoden 2008, 782.

ist, verhält es sich mit der Effektgröße genau umgekehrt: Mit zunehmender Stichprobengröße nähert sie sich ihrem wahren Wert in der Grundgesamtheit, weil man ja einen größeren Teil dieser Grundgesamtheit überblickt.[286] Während eine große Stichprobe mitunter nur deshalb signifikante Ergebnisse liefert, weil der hypothetische Null-Effekt nicht exakt null ist sondern minimal um null schwankt, wird ein Blick auf die Effektgröße diese Schwankung immer als das erkennen, was sie ist: minimal.

V. Ein Bild sagt mehr als tausend Signifikanzen.

Die fünfte Rezeptionsregel greift einen Themenkomplex auf, der bislang nicht angesprochen wurde, aber die vorangegangene Regel bestätigt und praktisch etwas handhabbarer macht.

Mathematische Präzision hat bekanntlich enorme psychologische Überzeugungskraft.[287] Deshalb wird empirischen Erkenntnissen oft umso mehr Glauben geschenkt, je komplizierter die zugrundeliegende Statistik ist und je feiner untergliedert die tabellarische Ergebnispräsentation – Quantifizierung gilt als Zeichen der Wissenschaftlichkeit:

> „Heute mehr denn je und vielleicht in den Vereinigten Staaten stärker als anderswo ist das Gefühl weit verbreitet, daß nur Zahlen Gewißheit geben. In dieser Beziehung herrscht tatsächlich eine Art volkstümlichen Aberglaubens vor [… der] gerade durch die Nüchternheit der statistischen Methoden – ein Zeichen ‚echter' Wissenschaft! – genährt wird. Aber selbst berufsmäßige Sozialwissenschaftler unterliegen ähnlichen Vorurteilen. Es gibt die Idee, daß Darstellungen sozialer Tatbestände unwissenschaftlich sind, wenn sie nicht von Zahlen gestützt werden. Die Folge ist ein ungezügelter ‚rage du nombre'."[288]

Diese „Zahlenwut" lässt sich vor allem durch die westliche Geistesgeschichte und ihre Betonung analytischer Rationalität erklären.[289] Demgegenüber galten bildliche Darstellungen wissenschaftlicher Erkenntnisse

---

[286] Deshalb gelten die vor allem in der Medizin praktizierten großangelegten RCT-Studien (*randomized controlled trial*) mit mehreren tausend Teilnehmern als "gold standard for estimating effects" (*Ellis*, Effect Sizes 2010, 116, 131 f.) und sind sogar Metastudien mit der gleichen zugrundeliegenden Stichprobengröße überlegen.

[287] *Beller*, Forschen lernen 2008, 77 f.; *Krämer*, Statistik 2011, 19: „Kleider machen Leute und Ziffern machen Zahlen, und je mehr Ziffern eine Zahl umhängen hat, desto mehr vertrauen wir ihr auch."; *Neumann/Krieger*, Clin L Rev 2003, 349, 382: "Precision can induce the illusion of completeness – the illusion that the precise number tells the whole story."

[288] *Nußbaum*, AcP 1955, 453, 472 f.; *Julian*, J Royal Stat Soc D 1987, 487, 489: "many people will only accept a measurement as scientific if it can be made to several significant figures on a scale."

[289] *Smith/Best u.a.*, Am Psy'st 2002, 749, 759: "With its logocentric bias, Western culture has traditionally favored the use of words, logic, and the formalisms of mathematics over the use of visual images, which have long been suspected of misleading the mind by appealing to the senses rather than to reason."

von jeher als provisorische Krücken für statistische Analphabeten und wissenschaftliche „Dilettanten“,[290] obwohl sie der biologisch evolvierten Wahrnehmungsfähigkeit des Menschen womöglich besser entsprechen als Zahlen.[291] Diese Einsicht konnte sich erst mit der Emanzipation der empirischen von den analytischen Wissenschaften allmählich durchsetzen.[292]

Heute bevorzugen gerade die Disziplinen, die als wissenschaftlichste wahrgenommen werden, Grafiken gegenüber Tabellen. So ist in den „exakten Wissenschaften“ Chemie, Physik, Biologie und Medizin der flächenmäßige Anteil, den Grafiken in Veröffentlichungen erhalten, mit durchschnittlich 14 % fast fünf Mal so hoch wie der mit 3 % in den Disziplinen Soziologie, Volkswirtschaftslehre und Psychologie:[293]

> „Je ‚härter‘ die Fakten, desto häufiger werden Grafiken und desto weniger häufig werden Signifikanztests verwendet. Das gilt sowohl für die Wissenschaften insgesamt als auch für die Teildisziplinen der Psychologie.“[294]

Während die als „hart“ wahrgenommenen Disziplinen bis zu zehn Mal mehr Platz für Grafiken als für Tabellen verwenden, gehen die „weichen“ Disziplinen genau umgekehrt vor.[295] Das war nicht immer so; viele der frühen psychologischen Erkenntnisse wurden mittels grafischer Auswertung statt schließender Statistik gewonnen.[296] Selbst einer der Gründungsväter der modernen Statistik nannte in seiner Vorlesung über „Geometrie der Statistik“ die grafische Auswertung „eine grundlegende Methode, um statistisches Material zu sichten und auszuwerten“, erkannte aber auch schon den „Wettbewerb von Geometrie und Arithmetik als wissenschaftliche Werkzeuge, um physikalische und soziale Phänomene zu bewältigen“.[297] Diesen Wettbewerb gewann zunächst die Arithmetik, weil die ersten Com-

---

[290] So *Luke Howard* 1847, zit. in *Wainer/Velleman*, Ann Rev Psy 2001, 305, 306 und 308.

[291] *Julian*, J Royal Stat Soc D 1987, 487, 488: “Human beings are excellent at examining patterns and pictures – evolution in the natural world has taken care of the training for that. They are very bad at doing mathematics. It was not an accident that geometry came before algebra.”; ähnl. *Wainer/Velleman*, Ann Rev Psy 2001, 305, 316 m.w.N.

[292] Ausf. *Wainer/Velleman*, Ann Rev Psy 2001, 305, 306 ff.

[293] *Smith/Best u.a.*, Am Psy’st 2002, 749, 753 unter Berufung auf *Cleveland*, Am Stat’n 1984, 261, der die genauen Zahlen zwar nicht berichtet, sie aber vielleicht auf Nachfrage mitgeteilt hat; vgl. auch *Epstein/Martin/Schneider*, Vand L Rev 2006, 1811, 1817: “most scientists have ‘declared a war’ on tables – or, at least, have expressed a strong preference for graphical displays”.

[294] *Sedlmeier/Renkewitz*, Forschungsmethoden 2008, 783.

[295] *Smith/Best u.a.*, Am Psy’st 2002, 749, 755.

[296] *Smith/Best u.a.*, Am Psy’st 2002, 749, 759: “Historically, many of psychology’s most enduring and best stabilized findings—ones that still decorate the pages of its textbooks—have been cast in the form of graphical displays, depicting relationships arrived at without the use of inferential statistics.”; vgl. auch *Wainer/Velleman*, Ann Rev Psy 2001, 305, 312.

[297] *Karl Pearson*, zit. nach *Best/Smith/Stubbs*, Behav Proc 2001, 155, 156 bzw. 162.

puter schließende Statistik schneller bewältigten, als sie Grafiken erstellen konnten.[298] Dieses Ungleichgewicht hat sich mittlerweile überlebt,[299] so dass eine Überprüfung der überkommenen Praxis naheliegt.

Die vermeintliche Präzision tabellierter Zahlenkolonnen hat nicht nur nichts mit der Exaktheit oder Härte einer empirischen Disziplin zu tun, sondern erschwert es mitunter sogar, „augen"fällige Unterschiede und Tendenzen auszumachen.[300] Das wohl anschaulichste Beispiel dafür wurde (nach seinem Erfinder) als *Anscombes Quartett* bekannt:[301]

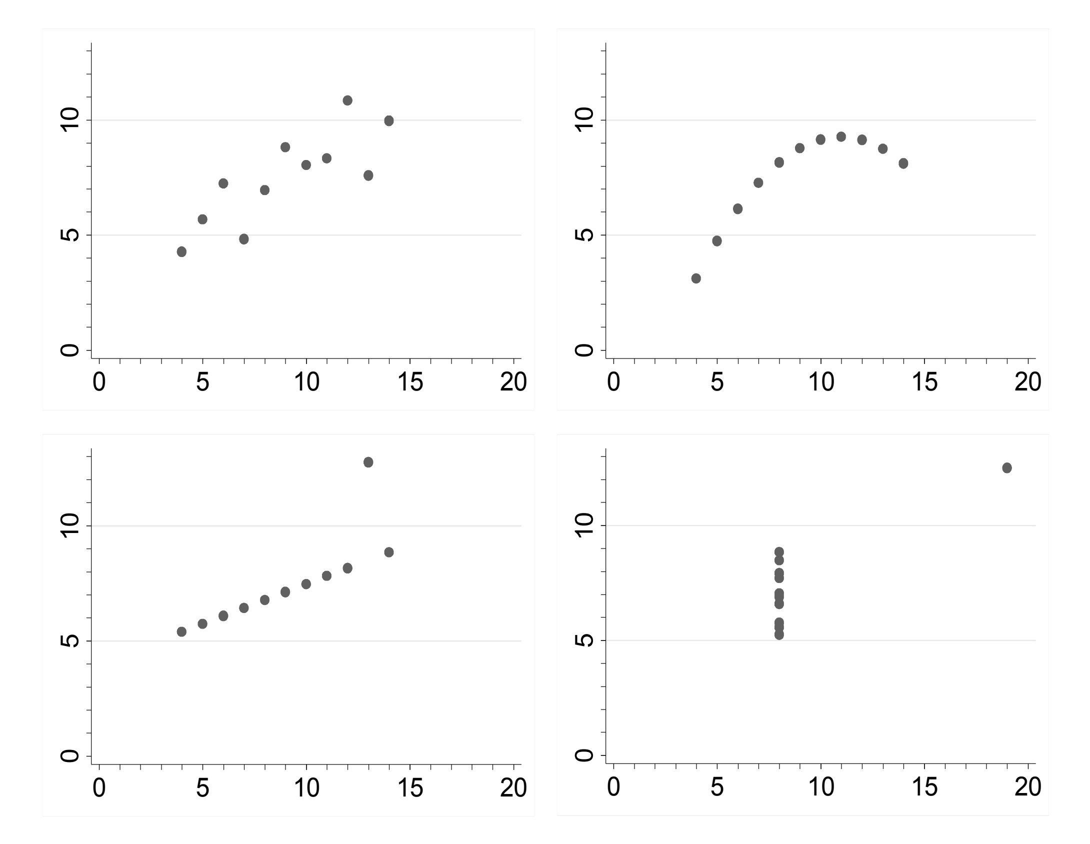


Illustration: Grafische Darstellung vier fiktiver Datensätze (eigene Abbildung)

[298] *Loftus*, Behav Res Meth 1993, 250, 250 f.; *Best/Smith/Stubbs*, Behav Proc 2001, 155, 156: "the period 1860–1890 has been called the 'golden age of graphs'. Many of the inferential statistics in use today were developed because of these early attempts to analyse data visually."

[299] *Wainer/Velleman*, Ann Rev Psy 2001, 305, 314 bezeichnen die auf heutigen Computern verfügbare bessere Software, größere Rechenleistung und größere Datenmenge als "the scientific triumvirate."

[300] *Smith/Best u.a.*, Am Psy'st 2002, 749, 753: "experimental research on human perception suggests that tables are inferior to graphs for conveying trends in data" m.w.N.

[301] *Anscombe*, Am Stat'n 1973, 17; dazu auch *Lawless/Robbennolt/Ulen*, Methods 2010, 294 f.; *Epstein/Martin/Schneider*, Vand L Rev 2006, 1811, 1845 f.

Die vier abgebildeten Datensätze folgen offensichtlich ganz unterschiedlichen Gesetzmäßigkeiten,[302] doch betrachtet man die herkömmlichen statistischen Parameter, stellt man fest, dass jeder einzelne dieser vier Datensätze exakt folgende Kennwerte aufweist:

N	= 11	(Stichprobengröße)
$\bar{X}$	= 9,000	(Mittelwert der unabhängigen Variable)
s_x	= 3,32	(Standardabweichung der unabhängigen Variable)
$\bar{Y}$	= 7,500	(Mittelwert der abhängigen Variable)
s_y	= 2,03	(Standardabweichung der abhängigen Variable)
r	= 0,816	(Korrelationskoeffizient)
b	= 0,500	(Regressionskoeffizient)
SE	= 0,118	(Standardfehler)
p	= 0,002	(Signifikanzwert)
R^2_{adj}	= 0,629	(adjustiertes Bestimmtheitsmaß)
$F_{(1,9)}$	= 18,0	(Modelltest mit neun Freiheitsgraden)

Illustration: Standardparameter der beschreibenden Statistik für die vier Datensätze[303]

Eine „blinde“ statistische Auswertung würde also zu dem Ergebnis kommen, dass in jedem der vier Datensätze eine Änderung der unabhängigen Variable um 1,000 zu einer gleichgerichteten Änderung der abhängigen Variable um 0,816 führt und dass in allen vier Datensätzen das gleiche lineare Abhängigkeitsverhältnis zwischen den beiden Variablen besteht ($y = 0,5\ x + 3$).[304] Dieses Ergebnis wäre nach allen Konventionen *signifikant*, also veröffentlichungsreif.

Nicht wenige Forscher plädieren deshalb – vor allem seit Pionierarbeiten über die erkundende Datenauswertung mittels grafischer Werkzeuge[305] sowie empirischen Vergleichsstudien zur Wahrnehmung von Tabellen und

---

[302] Das Bild links unten stellt die Funktion $y = 0,345x + 4$, das Bild rechts unten die Konstante $x = 8$ dar, jeweils mit einem Ausreißer. Im Bild rechts oben folgen alle Datenpunkte einer Funktion, die schon *Anscombe*, Am Stat’n 1973, 17, 19 als “possibly quadratic” bezeichnete: $y = -0,127x^2 + 2,781x - 6$.

[303] Rohdaten und Parameter nach *Anscombe*, Am Stat’n 1973, 17, 19 f.

[304] Diese vermeintlich überraschenden Übereinstimmungen sind eine praktische Konsequenz der oben (nach Fn. 101) beschriebenen statistischen „Abstraktion“ und bestätigen *van Belle*, Rules of Thumb 2008, 198: “Always graph the data […] There are any number of horror stories associated with ungraphed statistical analyses.”; *Lempert*, J Emp L Stud 2010, 907, 921: “Averages, even if accurately assessed, can hide lots of variance.”

[305] *Anscombe*, Am Stat’n 1973, 17, *Wainer/Thissen*, Ann Rev Psy 1981, 191, *Krohn*, Biol Phil 1991, 181 und *Wainer/Velleman*, Ann Rev Psy 2001, 305, 313 ff. zitieren als Ausgangspunkt die Arbeiten über “exploratory data analysis” von *John Tukey* – dem „Picasso der Statistik“, so *Salsburg*, Statistics 2001, 229 (Kap. 21).

Grafiken[306] – für einen vermehrten Einsatz von Grafiken als Ergänzung oder gar als völligen Ersatz für Auswertungen durch schließende Statistik (*visual inference*).[307] Neuere Studien verwenden bisweilen fast ausschließlich Grafiken, um die Ergebnisse fortgeschrittener statistischer Verfahren anschaulich zu vermitteln.[308]

Auch der Rezipient empirischer Forschung sollte deshalb nicht davor zurückschrecken, grafische Illustrationen aktiv nachzufragen und zu interpretieren.[309] Ein wohlbekanntes Sprichwort unter empirischen Forschern besagt, dass ein Ergebnis, das sich nicht in einer Grafik darstellen lässt, kein echtes Ergebnis sein kann.[310] Das steht vollkommen in Einklang mit der vorigen Rezeptionsregel, denn im Gegensatz zur Signifikanz lässt sich die Effektgröße immer schon aus der grafischen Darstellung ablesen, nämlich als die mit bloßem Auge erkennbare Differenz zwischen zwei Säulen oder Linien.[311]

Gleichwohl bergen natürlich auch grafische Darstellungen ein gewisses Irreführungspotential, das vor allem in pseudo- und populärwissenschaftlichen Grafikvarianten (*pop charts*) genutzt wird.[312] Dass grafische Darstellungen die formale Statistik vollständig entbehrlich machen, lässt sich daher mit guten Gründen bezweifeln.[313]

---

[306] *Gelman/Pasarica/Dodhia*, Am Stat'n 2002, 121: "The clearest conclusion from most of these studies is that tables are best suited for looking up specific information, and graphs are better for perceiving trends and making comparisons and predictions"; vgl. auch *Epstein/Martin/Schneider*, Vand L Rev 2006, 1811, 1822 bei Fn. 26.

[307] *Wainer/Thissen*, Ann Rev Psy 1981, 191; *Loftus*, Behav Res Meth 1993, 250; *Cumming*, Teach Stat 2007, 89; *Epstein/Martin/Schneider*, Vand L Rev 2006, 1811, 1838 ff. und *Epstein/Martin/Boyd*, Vand L Rev 2007, 801, 804 f.; philos. *Krohn*, Biol Phil 1991, 181.

[308] Musterbeispiel *Ho/Rubin*, Ann Rev L Soc Sci 2011, 17, die ihre gesamte statistische Auswertung in 19 Graphen wiedergeben – bei 18 Seiten Gesamtumfang – und nur für beschreibende Statistik zwei Tabellen vorhalten.

[309] Sehr hilfreiche Orientierungsregeln geben *Cumming/Finch*, Am Psy'st 2005, 170.

[310] Ähnl. *Senn*, Significance 2007, 79: "'If you need statistics to prove it, I don't believe it' is a remark you get used to hearing as a medical statistician"; zust. *Gorard*, Brit J Sociol Edu 2006, 67, 76 m.w.N.; a.A. *Julian*, J Royal Stat Soc D 1987, 487, 490.

[311] *Croson*, U Ill L Rev 2002, 921, 941: "apply the 'inter-ocular trauma test' […]; the differences between treatments should jump up and hit the reader in the eye."; nach *Epstein/Martin/Schneider*, Vand L Rev 2006, 1811, 1835 ff. machen gute Grafiken auch dadurch Signifikanztests entbehrlich, dass sie Wertestreuungen mit abbilden.

[312] Bsp. bei *Krämer*, Statistik 2011, 41 ff. (Kap. 3), 113 ff. (Kap. 9); *Epstein/Martin/Schneider*, Vand L Rev 2006, 1811, insb. 1823 Fn. 27.

[313] *Abelson*, Principled Argument 1995, 77: "it is easy to overreact to appearances in graphs."; *Salsburg*, Statistics 2001, 96: "Those things that seem to the eye to be similar […] are often drastically different when examined carefully with statistical tools developed for this purpose."; *Fisch*, Behav Proc 2001, 137: "Errors in trend detection illustrate the liabilities of using visual inspection as the sole means by which to analyze behavioral data."

VI. Statistik darf keine Magie, muss aber MAGIC sein.

Mehr noch als die Illusion der Präzision ist es vor allem die Unübersichtlichkeit der statistischen Methoden (vgl. oben vor Fn. 123), die empirische Studien bisweilen wie Zahlenzauberei erscheinen lässt.[314] Selbst hauptberuflichen Statistikern erscheinen neue Methoden mitunter so märchenhaft, dass für deren Benennung der Baron von Münchhausen Pate steht,[315] oder dass das dritte der drei Clarkeschen Gesetze auf den Plan gerufen wird:

„Jede hinreichend fortschrittliche Technologie ist von Magie nicht zu unterscheiden."[316]

Diese Metaphern sind selbstverständlich scherzhaft gemeint,[317] doch das englische Wort *magic* gewinnt einen durchaus ernsten Bezug zur Statistik, wenn man es als Akronym für fünf wichtige Faktoren versteht, denen die rhetorische Überzeugungskraft empirischer Argumente entspringt:[318]

1. Ausmaß der beobachteten Unterschiede (*Magnitude*), gemessen nicht an der Signifikanz, sondern an der Effektgröße,[319]
2. Klarheit der Darstellung (*Articulation*), hinsichtlich der Hauptaussagen (*ticks*) und ihrer Geltungsbedingungen und -grenzen (*buts*),[320]
3. Allgemeinheit der Erkenntnis (*Generality*), insbesondere aufgrund ihrer Bewährung in Variationen innerhalb oder zwischen Studien,[321]

---

[314] *Sedlmeier/Renkewitz*, Forschungsmethoden 2008, 781: „Der Signifikanztest […] hat in manchen Bereichen der Wissenschaft fast schon magischen Status erreicht."; *Vickers*, Understand Statistics 2010, 88: "many non-statisticians seem to think that [multivariable regression] has near magical properties."

[315] Ich meine die inzwischen anerkannte Methode von *Efron*, Ann Stat 1979, 1, Parameter *einer Grundgesamtheit* mittels Stichproben zu schätzen, die *aus einer Stichprobe* gezogen wurden – das *bootstrapping,* benannt nach der Geschichte, wie Münchhausen sich am eigenen Schopf (engl.: an den eigenen Stiefelschlaufen, *bootstraps*) aus dem Moor zog; Einführung bei *Sedlmeier/Renkewitz*, Forschungsmethoden 2008, 597 ff. (Kap. 19.1).

[316] Etwa Professor Sir *David Hendry* in einem Vortrag über "Empirical Model Discovery and Theory Evaluation" am 25.7.2011 in Jena; die Clarkeschen Gesetze sind nach dem Vorreiter der literarischen Wissenschaftsfantastik *Arthur Clarke* benannt, der sie 1973 aufstellte; Nachweise in *Prucher*, Science Fiction 2007, 22.

[317] Ironisch auch *Wainer*, Psy Meth 1999, 250, 255: "The magic of statistics cannot create information when there is none."; *Lempert*, J Emp L Stud 2010, 907, 909: "empirical studies are almost never a magic bullet."

[318] Zum Folgenden *Sedlmeier/Renkewitz*, Forschungsmethoden 2008, 780 f. m. Verw. auf *Abelson*, Principled Argument 1995, 11 ff.

[319] *Abelson*, Principled Argument 1995, 39 ff. (Kap. 3), auf 47 ff. mit dem Konzept einer *causal efficacy.*

[320] Näher *Abelson*, Principled Argument 1995, 104 ff. (Kap. 6), der die Begriffe prägte.

[321] *Abelson*, Principled Argument 1995, 132 ff. (Kap. 7).

4. Originalität der Erkenntnis (*Interestingness*), insbesondere aufgrund ihrer praktischen Bedeutung oder überraschenden Gewinnung,[322] und

5. Glaubwürdigkeit der Erkenntnis (*Credibility*), abhängig von den eingesetzten Methoden einerseits und der Kohärenz der resultierenden Theorie andererseits.[323]

Diese fünf Komponenten der Überzeugungskraft, von denen sich drei auf die Methoden beziehen und die beiden letzten auf den Inhalt des damit geführten Arguments, sind fast alle schon an anderer Stelle dieser Arbeit zur Sprache gekommen: Das Ausmaß (M) oben unter IV., die Allgemeinheit (G) unter dem Stichwort der „äußeren Gültigkeit" oben bei Fn. 44 sowie gleich in § 3 A.III., die Glaubwürdigkeit (C) hinsichtlich der Methoden unter dem Stichwort der „inneren Gültigkeit" oben bei Fn. 45 sowie gleich in § 3 A.II.[324] Das Klarheitskriterium (A) war bereits oben nach Fn. 283 in der Forderung nach präzisen Hypothesen angedeutet: Je präziser eine empirische Behauptung formuliert und eingegrenzt ist, desto leichter ist sie überprüf- und kritisierbar (sie ist „empirisch gehaltvoller") und desto bedeutsamer ist die Tatsache, dass sie bisher nicht widerlegt wurde.[325] Die ersten drei und der letzte Buchstabe des MAGIC-Akronyms fassen also vor allem die bereits erörterten Rezeptionskriterien zusammen. Neu ist demgegenüber das Kriterium der Originalität (I), das das Augenmerk wiederum auf die rhetorische Komponente empirischer Forschung lenkt und die Anschlussfähigkeit der juristischen Rezeption bekräftigt.

Insgesamt ist das MAGIC-Akronym also gut geeignet, die Rezeption empirischer Forschung anzuleiten und pragmatisch zu strukturieren. Dass eine bestimmte Studie *einem* der fünf Kriterien nicht genügt, ist sicher kein Ausschlusskriterium,[326] kann aber den Rezipienten dafür sensibilisieren, dass die Überzeugungskraft der mit ihr gewonnenen Erkenntnisse gemin-

---

[322] *Abelson*, Principled Argument 1995, 156 ff. (Kap. 8): "If a claim is so blah that no one cares to read or talk about it, the chances are small that it will enter the lore of a field—much less stimulate further investigation."

[323] *Abelson*, Principled Argument 1995, 170 ff. (Kap. 9); nach *Sedlmeier/Renkewitz*, Forschungsmethoden 2008, 781 auch bezogen auf glaubwürdige Hypothesenherleitung (dazu oben bei Fn. 271); aus juristischer Sicht *Ho/Rubin*, Ann Rev L Soc Sci 2011, 17, 28: "Credibility hence depends on deep, substantive knowledge of the legal system being examined."

[324] *Abelson*, Principled Argument 1995, 11 Fn. 5 sieht seine MAGIC sogar als Alternative zur Gültigkeitenlehre.

[325] *Hussy/Schreier/Echterhoff*, Forschungsmethoden 2010, 30; *Bortz/Döring*, Forschungsmethoden 2006, 4 ff.; vgl. auch *Gorard*, Brit J Sociol Edu 2006, 67, 76: "Clarity […] exposes our judgements to criticism […] Transparency does not, in itself, make a conclusion true or even believable, but it forces the analyst to admit the subjectivity of their analysis and allows others to follow their logic as far as it leads them."

[326] *Abelson*, Principled Argument 1995, 170 selbst postuliert "the *rule of two criticisms*" mit Erläuterung in Fn. 1.

dert sein dürfte. Umgekehrt garantiert aber auch die beste MAGIC-Umsetzung nicht die Überzeugungskraft der Studienerkenntnisse, denn Einzelstudien sind immer fehlerbehaftet. Das ist Gegenstand der siebten und letzten Rezeptionsregel.

VII. Synthese vor Primärstudien, Forschungsauswertung vor -bericht.

Die mittleren fünf Rezeptionsregeln betrafen im Wesentlichen die Rezeption einzelner Primärstudien. Einzelfälle sind das tägliche Brot der Juristenzunft. Schon in der juristischen Ausbildung ist *der Fall* die Standardeinheit, an der Wissen vermittelt und Erkenntnis gewonnen wird. Fälle sind anschaulich und leicht verdaulich und schon die Mutter der Jurisprudenz, die antike Rhetorik, kannte zahlreiche Stilmittel (Allegorien, Metonymien, Synekdochen, usw.) zur Veranschaulichung abstrakter Sachverhalte mittels lebensnaher Beispiele. „Juristen erläutern ihr Thema gerne anhand von Beispielen",[327] deshalb überrascht es nicht weiter, wenn Juristen auch bei der Rezeption empirischer Forschung oft besonderen Wert auf anschauliche und plausible Einzelstudien legen.[328] Dagegen ist zunächst auch nichts einzuwenden, denn empirische Forschung orientiert sich nicht zuletzt am Ideal des Geschichtenerzählers (vgl. oben bei Fn. 131), und ihre rhetorische Überzeugungskraft – das hat die vorige Rezeptionsregel ergeben – hängt maßgeblich von der Originalität ihrer Aussagen ab (MAG**I**C).

Zugleich beruht ihre rhetorische Überzeugungskraft aber auf Übertragbarkeit (MA**G**IC), die sich an einer einzelnen Studie niemals ablesen lässt.[329] Einzelne Studien sind zwangsläufig mit Unsicherheit behaftet und die Ergebnisse selbst der besten Studie können schlicht falsch sein.[330] Zumal oft zweifelhafte Kriterien verwendet werden, um die „besten" Studien zu identifizieren – etwa die Plausibilität der Ergebnisse (die letztlich immer die herrschende Weltsicht gegenüber neuen Erkenntnissen bevorzugt, wie in dem in § 4 vor Fn. 391 dargestellten Beispiel) oder die Angabe, wie oft eine Studie schon zitiert wurde (was wiederum vor allem die Plausibilität der Ergebnisse widerspiegelt). Der einzige Weg, *gute* Studien auszumachen – also solche, deren Fehlerwahrscheinlichkeit möglichst gering ist –,

---

[327] *Fleischer* in: Wank u.a., FS Wiedemann 2002, 827, 828.

[328] *Fleischer*, ZGR 2008, 185, 187 f. m.w.N. kritisiert das als Überbetonung von *legal narratives*; als Beispiel mag *Bainbridge*, Vand L Rev 2002, 1 dienen, der sich durchweg (12 f., 17, 22, 30, 39) auf ein "classic experiment" von 1932 stützt, nachdem zum gleichen Thema schon hunderte weiterer Studien durchgeführt wurden.

[329] So auch *Fleischer*, ZGR 2008, 185, 188 m.w.N.: „es bedarf vielmehr zusätzlicher Belege, um von einer *Einzelfallanalyse* zu einer statistisch oder empirisch gehaltvollen *Gesamtaussage* zu gelangen. Fehlen sie, bleiben alle Schlussfolgerungen Pseudo-Diagnostik."

[330] *Abelson*, Principled Argument 1995, 149: "One might say that isolated claims are not *robust*."

ist ein Blick auf die konkrete Durchführung und die Auswertungsmethoden. Und selbst dann lässt der Veröffentlichungsprozess (oben A.VI.) Zweifel daran, ob nicht zahlreiche *ebenso gute* Studien, die das Gegenteil belegen, unveröffentlicht geblieben sind.[331]

Bildlich gesprochen lässt sich das vielleicht so auf den Punkt bringen: Ein Jurist, der empirische Forschung aus einer einzelnen – noch so gut veröffentlichten, noch so viel zitierten und noch so plausiblen – Primärstudie rezipiert, ist wie ein Sozialwissenschaftler, der den aktuellen Rechtsstand aus einem viel beachteten Urteil in der NJW herleitet. Wo das hinführt, zeigt der hunderttausendfach als „Muster-Disclaimer" im Internet verbreitete Sinnspruch über die Verantwortung für verlinkte Webseiten,[332] der Internetseitenbetreiber seit Jahren in der trügerischen Sicherheit wiegt, für die Inhalte verlinkter Seiten nicht verantwortlich zu sein, solange sie sich nur deutlich von den verlinkten Inhalten „distanzieren" und ausdrücklich ihre Haftung ausschließen.[333] Mit dem Stand der Rechtsentwicklung hat das nichts zu tun.[334] Freilich: Ein Urteil mag das einzige sein und schon deshalb den Rechtsstand tatsächlich abbilden. Es mag das bestdurchdachte sein und deshalb zeitlosen Wert für die Rechtsfindung gewinnen. Und doch beruht es mit beträchtlicher Wahrscheinlichkeit mehr auf den zufälligen Besonderheiten des Einzelfalls als auf einer allgemeingültigen Schöpfung rechtlicher Prinzipien.

Nicht anders verhält es sich mit empirischen Studien: „Eine Studie ist nicht viel mehr als keine Studie, und die erste Studie zu einem neuen Verfahren unterliegt in aller Regel dem ‚Optimismus-Bias'".[335] Keine einzelne Studie kann die „großen Fragen" schlüssig beantworten,[336] deshalb lebt empirische Forschung von der Kumulation: „Eine einzelne Untersuchung ist

---

[331] *Krämer*, Denkste! 2011, 194: „Wir lesen nicht, wie viele andere Studien und Stichproben *ohne* signifikante Resultate es außerdem gegeben hat […] und ehe wir das nicht wissen, können wir auch die ‚Signifikanz' der selektierten Resultate, die uns letztendlich dann erreichen, nicht ermessen."

[332] In der frühen Fassung einer von mir selbst betriebenen Website lautete dieser Disclaimer: „Laut Urteil des Landgerichts Hamburg vom 12.5.1998 (AZ: 312 O 85/98) müssen Inhalte verlinkter Seiten mitverantwortet werden, wenn der Betreiber sich nicht ausdrücklich von den externen Inhalten distanziert."

[333] Der in der vorigen Fn. zitierte Disclaimer fuhr fort: „Der Autor […] distanziert sich daher ausdrücklich von jeglichen Seiten, auf die hier durch externe Links verwiesen wird. Er übernimmt keine Verantwortung für die Inhalte hier verlinkter Seiten und kann nicht für Schäden gleich welcher Art haftbar gemacht werden, die durch das Besuchen der hier verlinkten Seiten entstehen."

[334] Prägnant schon *Höcker*, Rechtsirrtümer 2007, 141 f.

[335] *Raspe*, GesR 2011, 449, 453.

[336] *Neumann/Krieger*, Clin L Rev 2003, 349, 386: "empiricism cannot answer – and nothing can answer – the big questions"; *Lawless/Robbennolt/Ulen*, Methods 2010, 47: "no individual study will be dispositive of a given research question."; *Lempert* (§ 2 Fn. 317).

günstigenfalls ein Mosaikstein. Erst viele Mosaiksteine formen ein Gesamtbild."[337] Erkenntnisfortschritt entsteht immer erst aus einer Synthese umfangreicher vorangegangener Forschung,[338] ebenso wie umgekehrt das Vorliegen einer Synthese darauf hinweist, dass ein Forschungsgebiet eine gewisse Maturität erreicht hat.[339] Diese Maturität bemisst sich vor allem an der Zahl durchgeführter Replikationen. Replikationen sind die sicherste Gewähr, die der empirische Forschungsprozess gegen Fehlurteile bietet (oben bei Fn. 194 und vor Fn. 273), daher sollten die Ergebnisse nicht replizierter Studien stets skeptisch behandelt werden.[340] Radikale Empiriker empfehlen sogar, die Ergebnisse einzelner Primärstudien niemals zu deuten – sondern lediglich so detailliert wie möglich zu berichten – denn Deutung sei erst möglich, wenn das Datenmaterial für eine Forschungssynthese ausreiche.[341] Die überwiegende Ansicht ist weniger restriktiv, aber ebenso deutlich: „Wenn Ergebnisse aus mehreren vergleichbaren Studien vorliegen, sind die zusammengefassten Schätzungen immer den Einzelschätzungen […] vorzuziehen."[342] Man wird also keinen ernsthaften empirischen Sozialforscher finden, der sich lieber auf eine solide durchgeführte Primärstudie stützt als auf eine solide durchgeführte Forschungssynthese.[343] Treffend wurde dieses Spannungsverhältnis zwischen Anschaulichkeit und Aggregation so zusammengefasst:

„Systematische Informationsgewinnung führt dazu, dass vorwiegend Fälle ausgewertet werden, die ganz gewöhnlich sind und keine Besonderheiten aufweisen. Wahrnehmungen von Praktikern scheinen dagegen dominiert von ungewöhnlichen Einzelfällen, die dann

---

[337] *Diekmann*, Sozialforschung 2012, 70; anschaulich spricht *Stanovich*, Psychology 2012, vom "connectivity principle" (124 ff.) und vom "principle of converging evidence" (128 ff.).

[338] *Allen/Preiss*, J Soc Behav Pers 1993, 9 zur "necessary and symbiotic relationship between meta-analysis and replication research"; *Nickerson*, Psy Meth 2000, 241, 291: "nothing of importance in psychology has ever been decided on the basis of the outcome of a single statistical significance test. Psychological knowledge is acquired, as is knowledge in other fields, as a consequence of the cumulative effect of many experiments and nonexperimental [sic] observations as well. It is the preponderance of evidence gathered from many sources and over an extended period of time that determines the degree of credibility".

[339] Vgl. etwa *Kozlowski/Ilgen*, Psy Sci Publ Interest 2006, 77, 79.

[340] *Allen/Preiss*, J Soc Behav Pers 1993, 9, "Scientific knowledge comes from replication. Nothing within the scientific community is, or should be, accepted until that finding becomes verified via multiple replications."; *Abelson*, Principled Argument 1995, 11: "Single Studies Are Not Definitive" und 77: "Research conclusions arise not from single studies alone, but from cumulative replication."; *Diekmann*, KZfSS Sonderheft 2004, 8, 27 ff. fordert eine „Replikationskultur".

[341] *Nickerson*, Psy Meth 2000, 241, 288 f. m.w.N.

[342] *Sedlmeier/Renkewitz*, Forschungsmethoden 2008, 778; *Lempert*, J Emp L Stud 2010, 907, 925: "Do not rest policy change or analysis on a single study, no matter how good it is."

[343] Vorausgesetzt, die Einzelstudie ist in der Synthese enthalten, sonst entscheidet die Stichprobengröße (vgl. Fn. 286).

in rechtspolitischen Diskussionen ein überdimensionales Gewicht bekommen. Schon die Juristenausbildung ist ja an ‚interessanten' Fällen ausgerichtet, die kein richtiges Bild abgeben von der täglichen Praxis in juristischen Berufen. Ähnlich verhalten sich alle Praktiker, die über ihre Arbeit berichten."[344]

Um diesen Gefahren zu entgehen, sollte der juristische Rezipient bei der Berücksichtigung empirischer Forschung immer Synthesen vor Primärforschung stellen und zuallererst gezielt nach Forschungssynthesen suchen. Nicht umsonst heißt es in der evidenzbasierten Medizin: „Originalstudien zu durchforsten, wo es gute aktuelle systematische Übersichten oder noch kompaktere kritische Aufbereitungen […] gibt, kann man […] als Zeitverschwendung einstufen."[345] Das schließt die Darstellung illustrativer Einzelstudien nicht aus,[346] solange sie das bleiben: illustrativ, nicht entscheidungstragend.

Der zweite Teil des Grundsatzes, der die beiden Hauptformen der Forschungssynthese ihrerseits in eine Rangordnung stellt, wird nicht ganz so einhellige Zustimmung finden. Schließlich ist nicht zu leugnen, dass die stärker subjektive Einfärbung des Forschungsberichts einige substantielle Vorteile mit sich bringt (dazu oben § 2 A.VII.1.). Ebenso wenig ist zu leugnen, dass auch die Forschungsauswertung subjektive Entscheidungen erfordert, die nicht nur offengelegt, sondern auch reflektiert werden müssen, damit Metastudien nicht die gleichen Mängel aufweisen wie viele Primärstudien – und allgemein jeder unreflektiert mechanische Einsatz statistischer Methoden.[347] Soweit ein Rezipient allerdings den Erkenntnisstand zu einer bestimmten Frage in Erfahrung bringen möchte und weniger an der wissenschaftshistorischen Entwicklung dieser Frage interessiert ist, sollte er die Forschungsauswertung stets gegenüber einem auf gleichem Niveau verfassten Forschungsbericht vorziehen (vgl. schon oben § 2 A.VII.2.c). Zu diesem Schluss kam auch die evidenzbasierte Medizin, in deren fünfstufiger Erkenntnishierarchie (*levels of evidence*) systematische Forschungsauswertungen die Spitze einnehmen.[348] Nicht umsonst konzentriert

---

[344] *Gessner* in: Plett/Ziegert, Empirische Rechtsforschung 1984, 69, 81 = *Gessner*, SozW 1984, 480, 484 f.; ähnl. *Heldrich*, AcP 1986, 74, 108: „Dabei werden die spektakulären Fälle mit atypischen Sachverhalten deutlich bevorzugt. Die Jurisprudenz des täglichen Lebens findet in der veröffentlichten Rechtsprechung kaum ihren Niederschlag."

[345] *Timmer*, DÄBl 2012, A 1418, 1420; vgl. auch *Howick/Chalmers u.a.*, OCEBM Levels 2011: "As always, a systematic review is generally better than an individual study."

[346] *Engel*, JITE 2010, 199, 202: "The facts that an experimental design has been colourful, that it resonates well with the policy problem at hand, that the key message is easy to grasp, all have to be factored in, and for good reason."

[347] *Ellis*, Effect Sizes 2010, 116: "Although meta-analysis has an aura of objectivity about it, in practice it is riddled with judgment calls. […] Unfortunately, many meta-analyses are done mechanically, with little attention given to these issues."

[348] *Howick/Chalmers u.a.*, OCEBM Levels 2011.

sich die evidenzbasierte Medizin seit ihren Anfängen vor allem auf die Metastudien,[349] die ein internationales Forschernetzwerk (*Cochrane Collaboration*) systematisch erarbeitet und regelmäßig aktualisiert.[350]

Auch die evidenzbasierte Jurisprudenz muss letztlich auf der in der Überschrift angedeuteten Erkenntnishierarchie beruhen: Das beste Abbild des Erkenntnisstands beinhalten Forschungsauswertungen, danach kommen Forschungsberichte und erst in letzter Linie, wenn keine der beiden Formen der Forschungssynthese aufzufinden ist, sollte sich ein Rezipient unmittelbar auf Primärforschung stützen.[351]

C. Weitere Literatur

Wie eingangs (vor A.) erwähnt, habe ich in den bisherigen Ausführungen auf Fußnoten möglichst verzichtet, um die Darstellung der gängigen Forschungsmethodik lesbar zu halten und nicht mit Nachweisen zu überfrachten. Zum Abschluss gehe ich daher auf weiterführende Lehrliteratur ein.

Eiligen Lesern sei ein kurzes, an Juristen gerichtetes Lehrbuchkapitel über „empirische Methoden“ nahegelegt, das der Jurist Petersen gemeinsam mit dem Wirtschaftswissenschaftler Goerg verfasst hat.[352] Auf vierzig leicht lesbaren Seiten mit anschaulichen Beispielen werden viele Schlüsselbegriffe erläutert, deren Systematik aber in der Kürze der Darstellung kaum zu erfassen ist. Ausführlichere Darstellungen finden sich in Lehrbüchern zur empirischen Methodik, von denen einige besonders empfehlenswert sind. Dabei stütze ich mich nicht auf eine systematische Durchsicht der gesamten Lehrbuchliteratur, sondern kann nur eine subjektiv gefärbte Auswahl derjenigen Bücher vorstellen, die ich selbst als juristischer Rezipient schätzen gelernt habe. Leser sind ausdrücklich eingeladen, weitere Lehrbücher aufzustöbern und zu verwenden, denn echtes Verständnis kann immer erst aus der Vervielfachung der Blickwinkel entstehen.

Im vorangegangenen Abschnitt habe ich vor allem sieben Lehrbücher systematisch ausgewertet – zwei große, ein mittleres und zwei kleine auf Deutsch sowie zwei weitere auf Englisch:

Das erste englische Lehrbuch, eine interdisziplinäre Gemeinschaftsarbeit des Juristen Lawless mit der Psychologin Robbennolt und dem Ökono-

---

[349] *Feinstein/Horwitz*, Am J Med 1997, 529, 530: “almost exclusive concentration on the ‘gold standard’ of randomized trials and meta-analyses”.

[350] Dazu *Bilger* in: Comberg/Klimm, Allgemeinmedizin 2004, 74, 75; ausf. *Higgins/Green*, Cochrane 2011.

[351] Das ist beispielsweise unten in § 4 C.III. der Fall, für die Beantwortung der Frage, ob Gruppen die Entscheidungsfindung entschleunigen.

[352] *Petersen/Goerg* in: Towfigh/Petersen, Methoden 2010, 201.

men Ulen, ist zugleich das bislang einzige Lehrbuch über „Empirische Methoden im Recht".[353] Es wendet sich an Rezipienten wie Produzenten gleichermaßen und leitet anschaulich Schritt für Schritt durch den Forschungsablauf. Kehrseite dieser narrativen Ausrichtung ist allerdings ein geringeres Maß an methodischer Systematisierung und Reflexion. Dafür endet jedes Kapitel mit Übungsaufgaben und das Buch mit einem Glossar.

Das große Lehrbuch von Bortz (mittlerweile verstorben) und Döring[354] wendet sich an alle „Human- und Sozialwissenschaftler" und folgt im Aufbau dem typischen Forschungsablauf. Es ist ein klassisch gehaltenes Standardwerk, das Konzepte klar und anschaulich erläutert – Glossar inklusive – und sich eng an die herrschende Methodik anlehnt, ohne ihre Mängel zu verschweigen.

Das große Lehrbuch von Sedlmeier und Renkewitz[355] richtet sich an Psychologen und schenkt dem Forschungs*ablauf* als solchem wenig Beachtung. Dafür legt es viel Wert auf Konzepte – das Buch beginnt mit Wissenschaftstheorie und endet mit einer Methodenreflexion – und vermittelt sogar die Grundzüge sehr moderner Auswertungsverfahren wie *bootstrapping* und bayesianischer Statistik.

Das mittelgroße Lehrbuch von Diekmann,[356] das schon in der 23. Auflage vorliegt, ist als „Enzyklopädie" bezeichnet, aber als Taschenbuch gedruckt. Und wie ein solches liest es sich, dank seiner flüssigen Prosa und zahlreicher unterhaltsamer Anekdoten. Der Autor ist Soziologie, lässt aber alle Disziplingrenzen hinter sich und orientiert seine übergreifende Darstellung der empirischen Methoden und statistischen Grundzüge konsequent am Forschungsablauf.

Das größere der beiden kleinen Lehrbücher von Hussy, Schreier und Echterhoff[357] ist an „Psychologie und Sozialwissenschaften für Bachelor" gerichtet. Nur die erste Hälfte betrifft überhaupt quantitative (der Rest qualitative und gemischte) Methoden, dafür reicht das Themenspektrum von „Was ist Wissenschaft?" bis „Datenerhebung im Internet", nur die Statistik kommt sehr kurz.

Das andere kleine Lehrbuch stammt von einem Psychologen (Beller),[358] richtet sich aber an keine bestimmte Disziplin. Es bietet eine überaus kompakte Darstellung (mit Beispielen und Arbeitsanregungen), die vielfach auf das große Lehrbuch von Bortz und Döring verweist und sich daher als Einstiegslektüre eignet. Vorbildlich ist die wiederholt betonte Differenzierung

---

[353] *Lawless/Robbennolt/Ulen*, Methods 2010.
[354] *Bortz/Döring*, Forschungsmethoden 2006.
[355] *Sedlmeier/Renkewitz*, Forschungsmethoden 2008.
[356] *Diekmann*, Sozialforschung 2012.
[357] *Hussy/Schreier/Echterhoff*, Forschungsmethoden 2010
[358] *Beller*, Forschen lernen 2008.

zwischen formal-forschungslogischen und inhaltlich-wertenden Entscheidungen.

Das siebte Lehrbuch schließlich, wiederum in englischer Sprache, stammt von Ellis[359] und umfasst drei Themen (Effektgrößen, Teststärken und Metastudien), die für die „Interpretation von Forschungserkenntnissen" und das Verständnis aktueller empirischer Entwicklungstendenzen unentbehrlich sind. Das Buch tritt mit dem Anspruch auf, seinen komplizierten Gegenstand „kauderwelschfrei" (*jargon-free*) zu präsentieren, und wird diesem Anspruch vollauf gerecht: Mit minimalen statistischen Vorkenntnissen flüssig lesbar, ebenso unterhaltsam wie fundiert, schließt es mit „30 Empfehlungen für Forscher" auf dem aktuellen Stand der Methodik.

Weiterhin habe ich mehrfach zitiert:

Drei kleine (weitgehend überschneidungsfreie) Taschenbücher über Wahrscheinlichkeit und Statistik von Krämer.[360] Teilweise seit mehr als zwanzig Jahren im Handel und kürzlich neu herausgegeben, sind sie populärwissenschaftliche „Bestseller". Obwohl für Laien geschrieben und anekdotisch statt systematisch geordnet, vermitteln sie auch dem wissenschaftlich-systematisch Interessierten, der ein gedankliches Ordnungssystem mitbringt,[361] solide und sehr plastische Einsichten in die Statistik.

Ein englischsprachiges Lehrbuch des Sozialpsychologen Abelson über „Statistik als prinzipiengeleitete Argumentation".[362] Während einige Abschnitte statistisches Vorwissen voraussetzen und einige der zahlreichen Beispiele veraltet erscheinen, verdeutlicht das Buch wie kein zweites den rhetorischen Charakter statistischer Datenauswertung und ist leicht lesbar geschrieben, mit viel Humor und einem seltenen Mut zur Originalität.[363]

Schließlich wurde einige empfehlenswerte Lehrliteratur vorliegend nicht zitiert:

Ein im wahrsten Sinne des Wortes anschauliches Lehrbuch über die Grundlagen der Statistik bietet die Gemeinschaftsarbeit des Statistikers Smith mit dem Grafiker Gonick.[364] Von den Anfängen der Wahrscheinlichkeitsrechnung bis hin zu statistischen Grundkonzepten bietet dieses Lehrbuch dem Statistikneuling eine kurzweilige und zügig lesbare Einführung im Comic-Format.

---

[359] *Ellis*, Effect Sizes 2010.

[360] *Krämer*, Statistik 2011; *Krämer*, Gebrauchsanweisung 2010; *Krämer*, Denkste! 2011.

[361] Zur beschreibenden Statistik *Krämer*, Gebrauchsanweisung 2010, 25 ff. (Kap. 2–4) sowie *Krämer*, Statistik 2011, 67 ff. (Kap. 5); zur schließenden Statistik *Krämer*, Denkste! 2011, 187 ff. (Kap. 8); zu Korrelation und Kausalität *Krämer*, Gebrauchsanweisung 2010, 183 (Kap. 13) sowie *Krämer*, Statistik 2011, 167 ff. (Kap. 14).

[362] *Abelson*, Principled Argument 1995.

[363] *Abelsons* Neuschöpfungen des MAGIC-Akronyms und der *ticks*- und *buts*-Terminologie (oben nach Fn. 318) sind nur die offensichtlichsten Beispiele.

[364] *Gonick/Smith*, Statistik 2009.

Mit einigem Vorwissen besonders gewinnbringend zu lesen ist das englische Lehrbuch über die „Grundlagen moderner statistischer Methoden“ von Wilcox.[365] Es richtet sich ausdrücklich an Nichtstatistiker und erläutert traditionelle Probleme der Statistik sowie „topaktuelle“ (*cutting edge*) Lösungsansätze, intuitiv und nahezu formelfrei. Es ignoriert Alternativen zur frequentistischen Statistik, erläutert diese dafür sehr eingängig und fundiert.

---

[365] *Wilcox*, Statistical Methods 2010.

§ 3

Die Gültigkeiten empirischer Forschung

„Wo eine Tatsache auf Grund wissenschaftlicher
Erkenntnis feststeht, ist für eine richterliche Feststellung
und Überzeugungsbildung naturgemäß kein Raum mehr.“[1]
– Bundesgerichtshof 1957

„Sowohl im allgemeinen praktischen als auch im juristischen Diskurs
[…] sind Regeln des vernünftigen Vermutens erforderlich. Dabei kann
die Relevanz empirischen Wissen für juristische Argumentationen
kaum überbewertet werden.“[2]
– Robert Alexy 1978

Das wichtigste Fazit des vorangegangenen Kapitels findet sich im siebten Rezeptionsgrundsatz,[3] demzufolge die juristische Rezeption sich in erster Linie auf Forschungssynthesen stützen sollte, weil nur sie den empirischen Erkenntnisfortschritt zuverlässig abbilden. Allerdings kommen juristische Rezipienten auf der Suche nach empirischer Erkenntnis um Primärstudien nie ganz herum, denn die Taktung der Rechtspolitik erlaubt es oft nicht, auf die für Forschungssynthesen erforderliche Evidenzmasse zu warten, zumal nie sicher ist, ob zu einer bestimmten rechtlichen Frage überhaupt je ausreichend empirische Forschung existieren wird.[4] Deshalb müssen juristische Rezipienten einzelne Primärstudien würdigen und ihre *Aussagekraft* beurteilen können. Aussagekraft, das wurde oben in § 2 A.II.1. bereits herausgearbeitet, bemisst sich an der Gültigkeit einer Studie – genauer: an *den* Gültigkeit*en* einer Studie, denn wie oben dargelegt lässt sich jede Studie anhand zweier unterschiedlicher, aber miteinander verbundener Gültigkeitsmaßstäbe beurteilen. Auf deren grundlegendes Spannungsverhältnis gehe ich zunächst näher ein (A.) bevor ich den Rest des Kapitels der Diskussion, Illustration und Erweiterung dieser Zweiteilung widme (B.).

---

[1] BGHSt 10, 208, 211, zit. nach *Upmeier*, Fakten 2010, 128; offen bleibt damit freilich, wann eine wissenschaftliche Erkenntnis als „feststehend“ angesehen werden kann.

[2] *Alexy*, Argumentation 1983, 287.

[3] Oben § 2 B.VII.

[4] Vgl. schon oben § 2 bei Fn. 67 und Fn. 208.

A. Das grundlegende Spannungsverhältnis

Die beiden Gültigkeitskriterien empirischer Forschung wurden oben (§ 2 A.II.1.) bereits eingeführt und sind nun noch genauer zu umreißen. Zuvor ist allerdings ein Vorbehalt nachzutragen, der bisher aus Gründen der Lesbarkeit unterschlagen wurde.

I. Vorbemerkungen zur Gültigkeitenlehre

Es ist natürlich vermessen, über „die beiden Gültigkeiten" zu sprechen, denn Gültigkeit ist keine ontologische Kategorie, sondern eine normative Setzung. Es gibt genauso wenig „zwei Gültigkeiten empirischer Forschung" wie es „zwei ethische Maßstäbe des Rechts" gibt. Doch auch wenn es nicht zwingend ist, Rechtsnormen zwischen den Polen der Nützlichkeits- und der Pflichtenethik zu verorten,[5] so hat es sich doch als nützlich und erkenntnisfördernd herausgestellt. Ebenso verhält es sich mit der Gültigkeitenlehre.

Wie die meisten wissenschaftlichen Lehren, ist sie eng mit ihrem Begründer verknüpft und letztlich nur einer von vielen möglichen Deutungsvorschlägen[6] – wenn auch derjenige, der sich kraft rhetorischer Überzeugungskraft und praktischer Bewährung am weitesten durchsetzen konnte. Er stammt von Donald Campbell, der ihn erstmals 1957 in Anlehnung an die psychologische Testtheorie formulierte, 1979 gemeinsam mit Thomas Cook die messtheoretischen Aspekte der Differenzierung noch stärker herausarbeitete, um schließlich die vormalige Zweiteilung vollends in eine Vierteilung aufzulösen.[7] Diese neue vierwertige Typologie enthält neben der inneren und äußeren Gültigkeit noch die statistische Schlussgültigkeit (*statistical conclusion validity*) sowie die Konstruktgültigkeit (*construct validity*).[8] Beide zu erläutern setzt allerdings soviel Vorwissen in Mess- und Testtheorie voraus, dass hier darauf verzichtet wird.

Ebenso verzichtet wird auf die Darstellung der vielen weiteren Gültigkeitskriterien,[9] die im Lauf der Zeit von anderen Autoren vorgeschlagen

---

[5] *Mastronardi*, Juristisches Denken 2003, Rn. 23: „Juristisches Denken bewegt sich zwischen den Polen der Nützlichkeits- und der Pflichtenethik (*Utilitarismus* und *Deontologie*)."

[6] Vgl. z.B. *Reichardt*, New Dir Eval 2011, 43 und oben § 2 Fn. 324.

[7] Zu dieser Entwicklung m.w.N. *Shadish/Cook/Campbell*, Causal Inference 2002, 37 ff.; *Heukelom*, J Econ Meth 2011, 13, 14 ff.

[8] Ausf. *Shadish/Cook/Campbell*, Causal Inference 2002, 42 ff. zur statistischen und 64 ff. zur Konstruktvalidität.

[9] Aus dem Stichwortverzeichnis von *Bortz/Döring*, Forschungsmethoden 2006: „Validität: differenzielle, diskriminante, externe, Inhalts-, interne, […] konvergente, Konstrukt-, Kriteriums-, prognostische […] statistische, Übereinstimmungs-" und dabei fehlen noch so populäre Schlagworte wie ökologische, diagnostische und Augenscheinvalidität (*face validity*).

wurden und einander oft überlappen oder aber ganz unvereinbare Grundpositionen reflektieren. Diese näher zu beleuchten wäre der Kohärenz und Verständlichkeit dieser Arbeit abträglich und verspräche auf ihrem Abstraktionsniveau kaum Mehrwert, denn die hier übernommene ursprüngliche Zweiteilung ist das leitende Ordnungsprinzip auch in der revidierten Campbellschen Gültigkeitenlehre[10] und ähnlichen Differenzierungen.[11] Daher belasse ich es bei der zweiwertigen Typologie, die Kennern der Materie unvollständig scheinen mag, für den hiesigen Zweck aber völlig ausreicht.

II. Eindeutigkeit (innere Gültigkeit)

Die innere Gültigkeit misst die Aussagekraft einer Studie hinsichtlich der eindeutigen Feststellung von Ursachenzusammenhängen.[12] „Messen" meint hier freilich nicht die Bezifferung mit einer Messgröße, sondern die normative Beurteilung, wie gut die Studie dem Zweck der Ein-Deutigkeit (oben § 2 nach Fn. 44) dient. Insofern lassen sich nur (aber immerhin) Kriterien dafür benennen, ob eine Studie empirisch Ursache und Wirkung unterscheiden kann. Es sind mindestens drei:[13]

Erstens müssen Veränderungen der vermeintlichen Ursache mit Veränderungen der vermeintlichen Wirkung einhergehen, und umgekehrt. Das wird *Kovariation* genannt (von lat. *com* und *variare*, zusammen verändern)[14] und lässt sich durch den aus der beschreibenden Statistik bekannten Korrelationskoeffizienten beziffern (dazu oben § 2 A.IV.1.). Die erste Voraussetzung lässt sich also auch so ausdrücken: Der Korrelationskoeffizient

---

[10] Ersichtlich in *Shadish/Cook/Campbell*, Causal Inference 2002 schon an der Kapitelgliederung ("2. Statistical Conclusion Validity and Internal Validity" ggü. "3. Construct Validity and External Validity") sowie ausdrücklich etwa auf 63 ("internal validity and statistical conclusion validity […] are closely related. Both are primarily concerned with study operations") und 93 ("Construct validity and external validity are related to each other […] both are generalizations").

[11] Heukelom, J Econ Meth 2011, 13, 17 f.: "An example is the 2000 eight-tome version of the APA's *Encyclopedia of Psychology*. The *Encyclopedia* contained two entries for validity, 'Validity', and 'Construct Validity' […] In addition to construct validity, validity was then divided into a number of categories and sub-categories. Only at the end did the validity-entry [… cite] internal and external validity as its main categories."

[12] Zu deren rechtlicher Relevanz eindringlich *Lempert*, J Emp L Stud 2010, 907, 926: "Always ask about mechanism; understanding why a situation exists is as important to policy analysis as knowing whether it exists."

[13] *Sedlmeier/Renkewitz*, Forschungsmethoden 2008, 127 ff. (Kap. 5.2.1); *Shadish/Cook/Campbell*, Causal Inference 2002, 6 f.; *Lawless/Robbennolt/Ulen*, Methods 2010, 31; *Henry*, Psy Inq 2008, 49, 66 m.w.N. setzt viertens "some logical mechanism" zwischen den Variablen voraus; *Thye* in: Webster/Sell, Experiments 2007, 60 ff. ersetzt das dritte Kriterium durch vier: *contiguity*, *nonspuriousness*, *consistency* und *theoretical plausibility*.

[14] *Dudenredaktion*, Fremdwörterbuch 2000, 731 („kon…"), 1388 („variieren"); *Harper*, Etymology 2014, Stichworte "com-", "variation" und "vary".

zwischen den Messgrößen für die vermeintliche Ursache und die vermeintliche Wirkung darf nicht null sein.

Zweitens müssen Änderungen der vermeintlichen Wirkung den Änderungen der vermeintlichen Ursache zeitlich nachfolgen. Das lässt sich mit einer Anleihe aus der Grammatik als *Nachzeitigkeit* bezeichnen.[15] Nachzeitigkeit ist zwar keine hinreichende Bedingung zur Feststellung von Kausalität – das entspräche dem logischen Fehlschluss des *post hoc ergo propter hoc* („danach, also deswegen“) – aber sie ist eine notwendige Bedingung. Die zweite Voraussetzung lässt sich also auch so ausdrücken: Ursachenrückschlüsse setzen Zeitreihendaten voraus. Sogar Experimente sind letztlich nur Zeitreihen aus gleichen Anfangszuständen und manipulierten Endzuständen.[16]

Drittens darf es *keine Verunreinigung* (*confound*)[17] der vermeintlichen Ursache geben.[18] Das heißt, dass keine Moderatorvariable (zum Begriff oben § 2 nach Fn. 68) ihrerseits mit der vermeintlichen Ursache stark kovariieren darf. Andernfalls wäre nicht zu klären, welche der beiden Variablen – die vermeintliche Ursache, die Moderatorvariable oder beide – die beobachtete Wirkung zeitigt. Dieses Problem ist aus der Rechtsdogmatik als „alternative Kausalität“ bekannt und liegt vor allem § 830 I 2 BGB, in gewissem Umfang aber auch § 231 StGB zugrunde.[19]

Je zweifelsfreier diese drei Voraussetzungen vorliegen, desto größer ist die Eindeutigkeit der Studie. Dabei bedarf es immer einer kumulativen Betrachtung der drei Voraussetzungen, denn auch Sozialwissenschaftler vernachlässigen oft, dass insbesondere die Korrelation eine zwar notwendige, aber keineswegs hinreichende Voraussetzung für einen empirischen Ursachenschluss ist.[20] Eine noch so starke Korrelation sagt nichts über die Ur-

---

[15] *Sedlmeier/Renkewitz*, Forschungsmethoden 2008, 128 sprechen umgekehrt von „zeitlicher Präzedenz“.

[16] So die Perspektive von *van Belle*, Rules of Thumb 2008, 58: “To establish cause and effect requires longitudinal data. […] Since causes precede effects it [sic] requires observations measured over time. […] The randomized experiment is essentially longitudinal. The randomization assures that the groups are comparable at baseline.”

[17] Von lat. *com* und *fundere*, zusammen schütten (*Harper*, Etymology 2014, Stichwort “confound”); vgl. *Dudenredaktion*, Fremdwörterbuch 2000, 736 („konfundieren“) und den verwandten Rechtsbegriff „Konfusion“: *Creifelds/Weber*, Rechtswörterbuch 2007, 682.

[18] *Ho/Rubin*, Ann Rev L Soc Sci 2011, 17, 22: “alternatively known as exogeneity, conditional exogeneity, ignorability, or selection on observables” – oder eben “impurities” (*Abelson*, Principled Argument 1995, 187 ff.).

[19] Vgl. *Wagner* in: Habersack, MK-BGB 2013, § 831 Rn. 28 ff., und zu § 231 StGB schon RGSt 9, 370, 380: „Schwierigkeit des Beweises [… ist] Grund der Aufstellung des Reates [Vorwurfs] des §. 227 [heute 231]“.

[20] *Abelson*, Principled Argument 1995, 181: “surprising how causal implications [.] sneak insidiously into interpretations of correlations.”; *Stanovich*, Psychology 2012, 80: “Tempting correlational evidence combined with a preexisting bias may deceive even the best of minds.”

sachenbeziehung aus, wenn nicht überzeugend dargetan werden kann, dass auch die Nachzeitigkeit vorliegt und Verunreinigungen fehlen.

III. Übertragbarkeit (äußere Gültigkeit)

Die äußere Gültigkeit misst die Aussagekraft einer Studie über den für die Forschungsfrage relevanten Lebenssachverhalt. Dabei wird herkömmlich von der „Verallgemeinerbarkeit" oder „Generalisierbarkeit" der Studie gesprochen (vgl. § 2 Fn. 44), doch ein genauerer Blick belegt zwei unterschiedliche Wege der Übertragbarkeit, von denen nur einer auf die aufsteigende Richtung (vom Speziellen zum Allgemeinen) beschränkt ist.

Dieser erste Weg ist zweifellos der vorzugswürdige, zugleich aber der weniger praktikable. Es ist derjenige der Verallgemeinerung aus der Zufallsstichprobe auf die Grundgesamtheit, also das oben (§ 2 A.IV.2.) eingeführte Grundprinzip der schließenden Statistik. Seine Stärke sollte bereits deutlich geworden sein: Er erlaubt den Rückschluss auf die Grundgesamtheit aus jedem ihrer Teile, weil ausgeschlossen werden kann, dass zwischen den Teilen systematische (d.h. nicht-zufällige) Unterschiede bestehen. Die Schwäche dieses Mechanismus ist seine begrenzte Anwendbarkeit.[21] Jede Studie hat mindestens vier Dimensionen, die mit UTOS abgekürzt werden: Untersuchungseinheiten bzw. Teilnehmer (*Units*), Studienbedingungen (*Treatments*), Zielgrößen (*Observations*) und Situationen (*Settings*).[22] Um wirklich verallgemeinerbar zu sein, müsste eine Studie auf jeder dieser Dimensionen echte Zufallsstichproben ziehen. Das ist nachgerade unmöglich. Selbst eine repräsentative Befragung von 100.000 deutschen Geschäftsführern über die Rolle von § 43 II GmbHG in ihrer täglichen Arbeit erlaubt streng genommen nur dann eine verallgemeinerbare Aussage über die Rolle von § 43 II GmbHG in der täglichen Arbeit deutscher Geschäftsführer, wenn die Stichprobe zufällig aus allen Geschäftsführern (U), aus allen möglichen Beziehungen (T) zur Geschäftsführerhaftung („in der Zeitung darüber gelesen haben", „es bei anderen beobachtet haben", „selbst erlebt haben"), aus allen möglichen Reaktionen (O) auf die jeweilige Beziehung („ignorieren", „extra provozieren", „gar nicht mehr aus dem Haus gehen") und aus allen denkbaren Aspekten der täglichen Arbeit eines Geschäftsführers (S) gezogen ist. Dabei ist unmittelbar einsichtig, dass in drei der vier Kategorien schon gar nicht klar ist, wie groß die Grundge-

---

[21] *Shadish/Cook/Campbell*, Causal Inference 2002, 24: "Formal sampling methods usually offer only a limited solution to causal generalization problems." (ähnl. 348); *Chui* in: McConville/Chui, Methods 2007, 46, 56: "randomisation is not always possible for various reasons".

[22] *Shadish/Cook/Campbell*, Causal Inference 2002, 19 m.w.N. zum Ursprung dieser Kategorisierung.

samtheit überhaupt sein könnte, und dass in der Teilnehmer-Dimension (U) nie eine Stichprobe aus allen Geschäftsführern möglich ist, sondern allenfalls aus den im Handelsregister eingetragenen, mit herkömmlichen Kommunikationsmitteln erreichbaren, die Frage verstehenden, des Schreibens fähigen, antwortwilligen und auch sonst kooperativen Geschäftsführern. Die Verallgemeinerung von der Stichprobe auf ihre Grundgesamtheit ist deshalb zwar das Ideal der empirischen Forschung, aber ein unerreichbares – gleich welche Art von Studie man unternimmt.[23]

Das für die Frage der äußeren Gültigkeit größere Gewicht kommt daher dem zweiten Weg zu, der von den Urhebern der Gültigkeitenlehre vorgeschlagen und als „Fundamentaltheorie übertragbarer Ursachenschlüsse" (*grounded theory of generalized causal inference*) beschrieben wurde.[24] Dabei handelt es sich um einen aus der Forschungspraxis abgeleiteten Katalog von fünf Kriterien, mittels derer die Übertragbarkeit einer Studie regelmäßig bewertet wird.[25] Nach dieser Übertragungslehre ist der Zielkontext, in den eine Erkenntnis übertragen werden soll, mit dem ursprünglichen Studienkontext qualitativ zu vergleichen,[26] und zwar mit Blick auf

1. ihre augenscheinlichen Ähnlichkeiten (*surface similarity*),
2. diejenigen ihrer Unterschiede, die als folgenlos bekannt sind (*ruling out irrelevancies*),
3. ihre belangvollen Unterschiede (*making discriminations*), die die Übertragung begrenzen,
4. die wechselseitige Hochrechenbarkeit (*interpolation and extrapolation*) ihrer UTOS-Dimensionen,[27] und

---

[23] *Mintken*, VR 1992, 252, 253: „Die zumeist begrenzten Untersuchungsmöglichkeiten schränken in der Regel auch den Übertragungsbereich merklich ein."

[24] *Shadish/Cook/Campbell*, Causal Inference 2002, 24 f., 348 ff.; *grounded theory* ist die gelegentlich (vgl. *Bortz/Döring*, Forschungsmethoden 2006, 332 ff.) als „gegenstandsbezogene Theoriebildung" übersetzte Bezeichnung für eine etablierte qualitative „Auswertungstechnik zur Entwicklung und Überprüfung von Theorien, die eng am vorgefundenen Material arbeitet bzw. in den Daten verankert (grounded) ist." (*Hussy/Schreier/Echterhoff*, Forschungsmethoden 2010, 194 ff.).

[25] So *Shadish/Cook/Campbell*, Causal Inference 2002, 348 f.; mit einer Analogie ließe sich daher vielleicht sagen, dass diese „Übertragungslehre" sich zur Verallgemeinerung aus Zufallsstichproben ungefähr so verhält wie *Kuhns* soziologische Lehre vom Paradigmenwechsel zu *Poppers* Wissenschaftstheorie vom inkrementellen Fortschritt durch Falsifikation und Bewährung (dazu *Sedlmeier/Renkewitz*, Forschungsmethoden 2008, 31 f.): Hier die Wissenschaftspraxis, dort das unerreichbare Idealbild.

[26] Ausf. zum Folgenden *Shadish/Cook/Campbell*, Causal Inference 2002, 353 ff.

[27] *Shadish/Cook/Campbell*, Causal Inference 2002, 366: "Both interpolation and extrapolation work best when more levels are sampled, when the functional form is well-identified over the sampled range, and when the extrapolation [sic] is to levels close to those originally sampled."

5. die Ähnlichkeit ihrer mutmaßlichen Ursachenstrukturen (*causal explanations*).

Je ähnlicher sich der Studien- und der Zielkontext in diesen fünf Dimensionen sind, desto größer ist die Übertragbarkeit von empirischen Erkenntnissen zwischen ihnen – nicht nur aus dem Speziellen ins Allgemeine, sondern auch andersherum, oder auf gleicher Ebene,[28] etwa von Berliner Komplementären auf Hamburger Komplementäre (Ortsebene), von Berliner Komplementären auf Berliner Kommanditisten (Sachebene) oder von Hamburger Kommanditisten 1990 auf Hamburger Kommanditisten heute (Zeitebene). Dabei ist es ratsam, einer Empfehlung der evidenzbasierten Medizin zu folgen und

> „bei der Frage nach der Übertragbarkeit [innerlich gültiger] Studienergebnisse auf einzelne […] Entscheidungssituationen nach einem Ausschlussprinzip zu verfahren. Danach wird die Übertragbarkeit so lange als gegeben angenommen, wie nicht begründete Argumente gegen eine solche Übertragung sprechen. Praktischer ausgedrückt wird also nicht die Frage beantwortet ‚Sind die Ergebnisse dieser Studie […] übertragbar?', sondern die Frage ‚Was spricht dagegen, dass die Ergebnisse dieser Studie […] übertragbar sind?'"[29]

B. Beobachtung und Experiment

Nachdem der normative Gehalt der beiden Gültigkeitskriterien nun näher bestimmt ist, geht es darum, ihre Anwendung im empirischen Forschungsalltag zu untersuchen. Wie oben (§ 2 A.II.1.) vorweggenommen, lassen sich Eindeutigkeit und Übertragbarkeit in eine methodische Zweiteilung von Beobachtung und Experiment übersetzen.[30]

Während Beobachtungen jeden Eingriff in die untersuchten Lebenssachverhalte idealerweise vermeiden,[31] sind solche Eingriffe im Experiment erforderlich, ja geradezu dessen Definitionsmerkmal. Daher stehen die Gültigkeiten empirischer Forschung regelmäßig im Widerspruch zueinander,

---

[28] Vgl. *Shadish/Cook/Campbell*, Causal Inference 2002, 21 f., 83 f.

[29] *Windeler*, Gesundheitswesen 2008, 418, 421.

[30] Ebenso *van Belle*, Rules of Thumb 2008, 3: "Broadly speaking there are two approaches to getting the data: observational studies and designed experiments."; *Arlen/Talley*, Experimental 2008, xv f.: "There are two generally accepted (and somewhat related) methodologies for testing theories of human behavior: (1) empiricism, which analyzes data drawn from the 'real world'; and (2) experimentalism, which analyzes data drawn from the laboratory."; *Epstein/King*, U Chi L Rev 2002, 1, 2; *Epstein/Martin* in: Cane/Kritzer, Hdb Empirical 2010, 901, 904.

[31] Beachte *Beller*, Forschen lernen 2008, 33: „Im Prinzip ist jede Datenerhebung eine Art Beobachtung. Im engeren Sinn […] allerdings nur das Sammeln von Daten in einem ‚nichtkommunikativen Prozess'."; vgl. auch *Bortz/Döring*, Forschungsmethoden 2006, 262 m.w.N.

und die Methodenwahl wird zur ersten und wichtigsten wertenden Entscheidung im Forschungsprozess.[32]

Die hier aufgegriffene Zweiteilung ist natürlich nicht die einzig denkbare (vgl. unten Fn. 229), erscheint mir aber besonders fruchtbar für die juristische Rezeption empirischer Forschung.[33] Im Folgenden stelle ich zunächst die beiden Idealtypen vor (II., III.) und erörtere sodann die Zwischenschattierungen (IV.). Denn wie sich zeigen wird, löst sich die Zweiteilung bei genauerem Blick in ein Spektrum auf, innerhalb dessen ganz verschiedene Gewichtungen von Übertragbarkeit und Eindeutigkeit möglich sind. Eingerahmt wird die Darstellung durch eine konkrete Rechtsfrage (I., V.), die die Darstellung der Studienarten illustriert.[34]

I. Zum Beispiel Anreizwirkung der Vorstandsvergütung

Die folgende Darstellung verschiedener Studienarten werde ich jeweils mit einer Beispielstudie verbinden, um die Möglichkeiten und Grenzen der jeweiligen Studienart zu illustrieren. Damit die Beispielstudien nicht unverbunden nebeneinander stehen, habe ich eine übergreifende Frage ausgewählt, zu der alle Beispielstudien, wenn auch mit ganz verschiedenen Methoden, beitragen:

Steigt die Leistung von Vorstandsmitgliedern mit höherer Erfolgsvergütung?

Diese Frage ist sowohl aus rechtsdogmatischer Sicht als auch aus rechtsökonomischer und rechtspolitischer Sicht relevant:

Rechtsdogmatisch schon deshalb, weil § 87 I 1 AktG den Aufsichtsrat verpflichtet, und zwar seit 2009 auch ausdrücklich,[35] „bei der Festsetzung der Gesamtbezüge des einzelnen Vorstandsmitglieds […] dafür zu sorgen, dass diese in einem angemessenen Verhältnis zu den […] Leistungen des Vorstandsmitglieds“ stehen.[36]

---

[32] *Shadish/Cook/Campbell*, Causal Inference 2002, 372: “Decisions that facilitate generalization will frequently come at the cost of […] other goals. For example, making irrelevancies heterogeneous […] at the cost of decreasing statistical power.”; vgl. unten bei Fn. 173.

[33] Sie ähnelt insb. der von *Mastronardi*, Juristisches Denken 2003, Rn. 35 ff. (Kap. II.1) dargestellten Dichotomie von Teilnahme und Beobachtung im juristischen Denken.

[34] Eingedenk *Hogarth*, J Acc Res 1982, S108, 109: “one does research for substantive reasons; thus, to evaluate research methodology without specifying either the nature of the specific phenomena considered, or the motivation for the particular research, is problematic”.

[35] Art. 1 Nr. 1 lit. a VorstAG (Gesetz zur Angemessenheit der Vorstandsvergütung) v. 31.7.2009, BGBl. I 2509; zur früheren Rechtslage, als § 87 I 1 AktG nur die „Aufgaben“, nicht aber die „Leistung“ erwähnte, vgl. nur *Thüsing*, ZGR 2003, 457; *Hoffmann-Becking*, ZHR 2005, 155, 158 f.; *Fonk*, NZG 2005, 248, 249 f.

[36] Dazu *Dauner-Lieb* in: Henssler/Strohn, GftsR 2014, § 87 AktG Rn. 16; *Fleischer* in: Spindler/Stilz, AktG 2010, § 87 Rn. 11 ff.; *Seibt* in: Lutter/Schmidt, AktG 2010, § 87 Rn. 9; rechtstatsächlich aus Praktikersicht *Ehren/Gros*, Konzern 2010, 412.

Dementsprechend stellt sich die Frage, welches Verhältnis von Bezügen und Leistung *angemessen* ist. Die objektiv-teleologische Auslegung dieses Begriffs ergibt, „dass Vergütung auch Anreize zur Leistungssteigerung setzen soll“,[37] daher ist weiter zu fragen, wann und warum die Vergütung leistungssteigernd wirkt. Das ist eine empirische Frage, und ihre Antwort ist nach § 116 S. 1, 3 i.V.m. §§ 93 II 1, 87 I 1 AktG für mindestens zwei Akteure relevant: Erstens für den Aufsichtsrat, der für die Erfüllung seiner gesetzlichen Pflicht haftet[38] – und deshalb etwa auf die Vereinbarung variabler Vergütungsbestandteile[39] angewiesen sein *könnte.*[40] Zweitens für die Richterin, die in den (eher theoretischen) Fällen, in denen Vergütungsvereinbarungen zum Gegenstand einer Haftungsklage werden, darüber befinden muss, ob der Aufsichtsrat *ex ante* davon ausgehen durfte, mit der gewählten Vergütungsregelung Leistungsanreize im Interesse des Unternehmens gesetzt zu haben.

Ganz ähnlich fragt sich aus *rechtsökonomischer* Sicht, ob die Vorstandsvergütung den Funktionen gerecht wird, die sich aus der mikroökonomischen Analyse des Rechts ergeben.[41] Kernelement dieser Analyse ist der rationale Eigennutzmaximierer (*homo oeconomicus*), der fremdnützige Arbeit zu vermeiden sucht, weil sie sowohl Arbeitsleid als auch Opportunitätskosten in Form entgangener Freizeit verursacht. Mindestens diese Kosten muss ein etwaiger Auftraggeber (*principal*) durch die Zahlung einer Vergütung kompensieren. Die einfachste Vergütungsform – der genau aufwandsgerechte Akkordlohn – lässt sich allerdings nur dort einsetzen, wo der Arbeitseinsatz des Auftragnehmers (*agent*) genau beobachtbar ist. Daran fehlt es bei nahezu allen Dienstleistungen, einschließlich jener, die der Vorstand für die Aktionäre (als Auftraggeber im Sinne der ökonomischen

---

[37] So *Dauner-Lieb* in: Henssler/Strohn, GftsR 2014, § 87 AktG Rn. 16; ebenso *Seibt* in: Lutter/Schmidt, AktG 2010, § 87 Rn. 1: „nicht nur Entgelt für geleistete Tätigkeit, sondern auch Instrument der Verhaltenssteuerung“; ebd. Rn. 8, 11 zum „funktionalen Kriterium des optimierten Leistungsanreizes und Steuerungseffektes“; *Ihrig/Wandt/Wittgens*, ZIP 2012, H. 40 Beil., 4 bei Fn. 23; krit. *Kort*, NJW 2005, 333, 335; ganz a.A. *Fonk*, NZG 2005, 248, 252.

[38] *Henssler* in: Henssler/Strohn, GftsR 2014, § 116 Rn. 25; *Spindler* in: Goette/Habersack, MK-AktG 2014, § 87 Rn. 126 ff.; *Seibt* in: Lutter/Schmidt, AktG 2010, § 87 Rn. 17; *Drygala*, ebd., § 116 Rn. 51 ff.

[39] So ohne Weiteres Ziff. 4.2.3 II 2 DCGK sowie bspw. *Dauner-Lieb* in: Henssler/Strohn, GftsR 2014, § 87 AktG Rn. 16; weitere Nachw. bei *Arnold*, Steuerung 2007, 122 Fn. 50.

[40] *Ihrig/Wandt/Wittgens*, ZIP 2012, H. 40 Beil., 12: „Ob bei einer Kombination verschiedener variabler Vergütungsbestandteile insgesamt ein längerfristiger Verhaltensanreiz gesetzt wird, ist […] in erster Linie auch eine tatsächliche Frage“; krit. *Bernhardt/Witt*, ZfB 1999, 825; *Arnold*, Steuerung 2007, 124 ff.; *Windbichler* in: Grundmann u.a., FS Juristische Fakultät 2010, 1079, 1093; psychologisch begründet *Hamann*, BRJ 2010, 27, 29 f. m.w.N.

[41] Neben der hier erläuterten gehören dazu v.a. die Risikoanreiz- und die Interessenkopplungsfunktion, vgl. *Schüller*, Vorstandsvergütung 2002, 46 f.; *Stenzel*, Vorstandsvergütung 2012, 94; krit. *Brauer*, NZG 2004, 502, 503.

Theorie) erbringt. Ist der Arbeitseinsatz der Vorstandsmitglieder allerdings unbeobachtbar (*hidden action*), so können sie ihren Nutzen dadurch maximieren, dass sie ihren Arbeitseinsatz minimieren (*shirking*) und dadurch *sowohl* die vereinbarte Vergütung *als auch* Freizeit gewinnen. Dieser Versuchung (*moral hazard*) zur Ausnutzung der Informationsasymmetrie können die Aktionäre nicht vertraglich entgegenwirken, daher sind sie auf die zweitbeste Lösung angewiesen (*second best*), den Vorstand von vornherein in gewissem Umfang erfolgsabhängig zu entlohnen, damit er schon im eigenen Interesse seinen Arbeitseinsatz maximiert.[42] Ob er das aber tatsächlich tut – ob also die Vergütung zu einer höheren Leistung führt – ist wiederum eine empirische Frage.

Schließlich sind empirische Erkenntnisse zu dieser Frage auch *rechtspolitisch* relevant. Die Vorstandsvergütung gehört zu den Dauerthemen von Rechtspolitik und Gesetzgebung seit 1937 und war Gegenstand ganz unterschiedlicher Regelungskonzepte.[43] Schon die Ermöglichung der optionsbasierten Vorstandsvergütung (*stock option plan*) durch § 192 II Nr. 3 AktG sollte dabei helfen, „das Management [.] zu einer an der langfristigen Wertsteigerung orientierten Unternehmensstrategie motivieren“.[44] Auch die heute geltenden Vorgaben zur Vergütungsstruktur in § 87 I 2, 3 AktG:

> „Die Vergütungsstruktur ist bei börsennotierten Gesellschaften auf eine nachhaltige Unternehmensentwicklung auszurichten. Variable Vergütungsbestandteile sollen daher eine mehrjährige Bemessungsgrundlage haben […]“

beruhen auf einer ganz bestimmten Vorstellung über die Anreizwirkungen der Vorstandsvergütung – nämlich „dass von kurzfristig ausgerichteten Vergütungsinstrumenten fehlerhafte Verhaltensanreize ausgehen können“ und dass deshalb etwa „Gratifikationen und Boni nicht so angelegt sein sollen, dass die Erfüllung ihrer Parameter nur zum Stichtag […] angestrebt wird“ oder dass „Aktien, die als Vergütung gewährt werden, Haltefristen unterliegen sollten, die sich an der Ausübungsfrist für Aktienoptionen orientieren“.[45] Ähnliches gilt für die unverbindliche, aber als „anerkannter Standard guter und verantwortungsvoller Unternehmensführung“ (Präambel DCGK) ausgegebene Empfehlung 4.2.3 des Deutschen Corporate Governance Kodex, dass „die monetären Vergütungsteile [.] fixe und variable Bestandteile umfassen“ sollen. Die diesen Vorgaben zugrundeliegenden

---

[42] Krit. *Bernhardt/Witt*, ZfB 1999, 825, 838: „In der Realität liegen die Dinge aber komplizierter.“

[43] *Hamann*, BRJ 2010, 27 f. m.w.N. identifiziert fünf historische Phasen mit je anderem Grundverständnis.

[44] Begr. Art. 1 Nr. 23 KonTraG, BT-Drs. 13/9712 v. 28.1.1998, 23.

[45] Begr. AT und Art. 1 VorstAG-RegE, BT-Drs. 16/12278 v. 17.3.2009, 5.

Annahmen müssen sich empirischer Überprüfung stellen und profitieren von einem genaueren Verständnis der Rechtswirklichkeit,[46] ja mehr noch:

> „Mit der Neufassung des § 87 Abs. 1 AktG hat der Gesetzgeber mittelbar *den Auftrag erteilt*, Anreizeffekte zu erforschen, soll die Rechtsordnung nicht auf dem Stand ungesicherter Alltagstheorien verharren."[47]

Das juristische Schrifttum ist sich durchaus bewusst, dass „die Grundannahme, durch moderne Vergütungsformen ließe sich […] eine größere Anreizwirkung auf das Management ausüben", bislang „nicht praktisch nachgewiesen" ist.[48] Die vorliegende Arbeit kann und will nur beschränkt zu diesem praktischen „Nachweis" beitragen, denn erstens liegen bereits Monographien vor, die auch auf empirische Erkenntnisse zurückgreifen (wenn auch noch nicht allzu methodenkritisch),[49] und zweitens müssten viele Facetten der Vergütungsgestaltung berücksichtigt werden (oben angeklungen sind bereits die Bedeutung langfristiger Anreizhorizonte und die Haltefristen für Belegschaftsaktien), um für eine gegebene Situation oder Situationstypen die jeweils anreizstärkste Gestaltung empirisch zu ermitteln.[50]

Um den Rahmen vorliegend nicht zu sprengen, beschränke ich mich auf die schon sehr konkrete (und für eine allgemeine Vergütungsevaluation *zu* konkrete) Ausgangsfrage, welchen Einfluss die Höhe einer erfolgsabhängigen Barvergütung auf die Leistungen des Vorstands hat. Zu dieser Frage wurden Studien mit ganz verschiedenen Methoden durchgeführt und sogar ein „evidenzbasierter Forschungsüberblick" vorgelegt,[51] daher halte ich sie für höchst anschaulich, um unterschiedliche empirische Herangehensweisen zu verstehen und zu erkennen, was jede von diesen für Übertragbarkeit und Eindeutigkeit der Erkenntnisse bedeutet.

---

[46] Krit. *Ihrig/Wandt/Wittgens*, ZIP 2012, H. 40 Beil., 10: „Ob die Kodexempfehlungen an dieser Stelle tatsächlich ‚gute' Corporate Governance reflektieren, ist zweifelhaft"; ähnl. *Gosling*, PwC 2012: „If executive pay were genuinely motivating executives towards higher levels of performance, with benefits for all, there would surely be less controversy about the subject. But is it?"

[47] *Windbichler* in: Grundmann u.a., FS Juristische Fakultät 2010, 1079, 1094 (Hervorhebung nur hier).

[48] *Schüller*, Vorstandsvergütung 2002, 86, die aber trotzdem ebd. 89 ohne Begründung unterstellt, „dass moderne Vergütungsformen und insbesondere Aktienoptionen [.] wohl einen positiven Einfluss auf die Leistungsbereitschaft von […] Vorstandsmitgliedern haben".

[49] *Schüller*, Vorstandsvergütung 2002, 85 ff.; *Stenzel*, Vorstandsvergütung 2012, 223 ff.

[50] Ausf. *Ihrig/Wandt/Wittgens*, ZIP 2012, H. 40 Beil., 10 ff.

[51] *Jacquart/Armstrong*, Interf 2013, 580; pessimistisch dazu *Pfeffer*, Interf 2013, 578: "I am certain that Jacquart and Armstrong's sensible, evidence-based suggestions for reforming executive pay have little chance of being implemented; they will face vigorous opposition from the various interests so well served by the present arrangements."

II. Korrelationsstudien – Idealtyp der Übertragbarkeit

Die größtmögliche Übertragbarkeit können Studien erzielen, die aufgezeichnete Daten aus der bereits vergangenen Lebenswirklichkeit verwenden. Da solche Studien ihre Hypothesen mittels schließender Statistik an bereits existenten Daten prüfen, lassen sie sich als *sekundär*statistische Korrelationsstudien bezeichnen.[52] Diese Korrelationsstudien sind gewissermaßen die reinste Form der Beobachtung, weil ihre Daten bereits existieren, die Studie also die untersuchten Lebensvorgänge auf keinen Fall beeinflusst haben kann. Dementsprechend weisen solche Studien die höchstmögliche äußere Gültigkeit auf: Sie verwenden unverfälschte Daten aus genau dem Lebenssachverhalt, auf den ihre Fragestellung bezogen ist. Studien- und Zielkontext decken sich, daher sind Korrelationsstudien nach der Gültigkeitenlehre (oben A.) der Idealtyp der Übertragbarkeit.

1. Ein Beispiel: Kaplan & Rauh (2010)

Eine vielbeachtete Studie zweier Ökonomen widmete sich der Frage, ob die Vorstandsgehälter in neuerer Zeit tatsächlich ungewöhnlich schnell gestiegen sind.[53] In diesem Zusammenhang werteten die Autoren auch aus, ob höhere Vorstandsvergütungen mit höheren Leistungen einhergehen.[54] Die Vorstandsvergütungen wurden der Datenbank „ExecuComp“ entnommen,[55] die alle 1.500 Unternehmen in den drei Hauptindizes der Ratingagentur Standard & Poor’s umfasst (S&P 500, S&P 400 mid-cap und S&P 600 small-cap) und damit 88 % der gesamten Marktkapitalisierung US-amerikanischer Aktiengesellschaften abbildet.[56] Die Leistungen eines Vorstands wurden anhand der um Brancheneffekte bereinigten Kapitalrendite des jeweiligen Unternehmens gemessen,[57] ähnlich wie in einer klassischen

---

[52] Zur Korrelation oben § 2 nach Fn. 111; der Begriff „Korrelationsstudie“ wird andernorts auch i.w.S. verwendet, etwa für jegliche Beobachtungsstudien: *Huber*, Experiment 2005, 70; *Hussy/Schreier/Echterhoff*, Forschungsmethoden 2010, 148 ff.

[53] *Kaplan/Rauh*, Rev Fin Stud 2010, 1004; Synopse bei *Bebchuk/Weisbach*, Rev Fin Stud 2010, 939, 946: “examine the question of whether the growth of pay can reflect market forces in a creative way. The idea is that if executive pay reflects market forces, then its growth should parallel that of other highly paid professions […] such as financial service sector employees from investment banks, hedge funds, private equity funds, and mutual funds (Wall Street), as well as corporate lawyers, professional athletes, and celebrities.”

[54] *Kaplan/Rauh*, Rev Fin Stud 2010, 1004, 1043 unter “6. Pay-for-Performance”; Zeitraum der Studie ist 1999–2004.

[55] *Kaplan/Rauh*, Rev Fin Stud 2010, 1004, 1005 bzw. 1044.

[56] *Cadman/Klasa/Matsunaga*, Acc Rev 2010, 1511.

[57] *Kaplan/Rauh*, Rev Fin Stud 2010, 1004, 1044: “total return to the firm’s stock less the value-weighted performance of the firm’s industry (using Fama-French 1997 [.] returns).”

Korrelationsstudie zwanzig Jahre zuvor.[58] Weitere Einzelheiten der Operationalisierung erübrigen sich für die Illustration der Herangehensweise; eine ausführliche Rezeption dieser Studie müsste freilich die Beschränkungen und normativen Konsequenzen der gewählten Operationalisierungen von „Vorstandsvergütung“ und „Leistung“ kritisch reflektieren.

Die Studie teilt alle betrachteten Unternehmen zunächst in fünf Größenkategorien (Quintile) auf, um der Tatsache Rechnung zu tragen, dass die Vorstandsvergütung mit der Unternehmensgröße steigt.[59] Innerhalb jedes Größenquintils ordnet sie sodann auch die Höhe der Vorstandsvergütungen in fünf Quintile und stellt fest, dass über alle Unternehmensgrößen hinweg steigende Vorstandsvergütungen mit steigender Kapitalrendite einhergehen:[60] Das bestbezahlte Fünftel der Vorstände erwirtschaftet eine durchschnittliche Mehrrendite von 61 % im Vergleich zum jeweiligen Branchendurchschnitt, während das schlechtestbezahlte Fünftel der Vorstände durchschnittliche Minderrenditen von 19 % erwirtschaftet.[61] Daraus schließen die Autoren, dass Vergütung und Leistung stark zusammenhängen.[62]

2. Methodische Beschränkungen

Da hier keine ausführliche Kritik an der Studie von Kaplan und Rauh, sondern die Illustration der Korrelationsstudie als einer empirischen Herangehensweise beabsichtigt ist, sind nunmehr ganz allgemein die methodischen Beschränkungen dieser Herangehensweise herauszuarbeiten.

Dabei ist zunächst zu erinnern, was die Korrelationsstudie auszeichnet: Sie ist die Reinform der empirischen Beobachtung. Der Forscher beobachtet genau den Lebenssachverhalt, der ihn interessiert, und ist zugleich frei von jedem Verdacht, diesen Lebenssachverhalt beeinflusst zu haben. Andererseits verhindert dieses Vorgehen jede Aussage über Ursachenzusammenhänge,[63] was sich vor allem in vier Beschränkungen äußert:[64]

---

[58] *Jensen/Murphy*, J Polit Econ 1990, 225, 227 berechneten eine “pay-performance sensitivity” als Verhältnis des marginalen Mehreinkommens von Vorständen zur marginalen Mehrrendite von Aktionären.

[59] *Kaplan/Rauh*, Rev Fin Stud 2010, 1004, 1044 m.w.N.

[60] Berichtet werden nur Dreijahresdurchschnitte, aber vgl. *Kaplan/Rauh*, Rev Fin Stud 2010, 1004, 1045: “The results are qualitatively identical for one year and five years of stock performance.”

[61] *Kaplan/Rauh*, Rev Fin Stud 2010, 1004, 1045 mit Trendgrafik auf 1044.

[62] *Kaplan/Rauh*, Rev Fin Stud 2010, 1004, 1044: “actual compensation is highly related to performance, i.e., there is strong pay-for-performance.”; zu den andernorts geäußerten rechts*politischen* Schlussfolgerungen des Studienerstautors vgl. *Bogle*, Acad Mgmt Persp 2008, 21.

[63] Vgl. *Klick*, JITE 2010, 166, 167: “researchers need to remain appropriately humble. The best a researcher can do in this area is to present results as purely descriptive, abandoning the search for causality and strong empirical tests of hypotheses.”

a) Endogenität (reverse causality)

Die erste Beschränkung ist diejenige der Endogenität. Das lässt sich sinngemäß mit „Ausgeburt“ übersetzen (oben § 2 Fn. 67) und bedeutet, dass die vermeintliche Ursache ihrerseits eine Ausgeburt (Wirkung) anderer Ursachen ist, sich also nicht in einer Messgröße isolieren lässt. Etwa könnte die vermeintliche Ursache in Wirklichkeit ganz oder teilweise durch die Messgröße verursacht worden sein, die man für ihre Wirkung hielt, so dass die Ursachenbeziehung umgekehrt von der vorgestellten verläuft (*reverse causality*)[65] – so finden sich in der Literatur sowohl Studien, die die Vorstandsleistung als abhängig von der Vorstandsvergütung betrachten, als auch umgekehrt.[66] Ebenso gut könnten aber beide Messgrößen durch eine dritte verursacht werden – dann spricht man von einer Bastardkorrelation (*spurious correlation*)[67] – wie in der berühmten Korrelation von Storchenpopulationen und Geburtenraten, die beide mit dem Urbanisierungsgrad steigen und fallen,[68] oder der in den USA kolportierten Korrelation zwischen Eisverkäufen und Mordraten, die beide von Jahreszeit und Wetter abhängen.[69] Endogenität in all ihren Ausprägungen ist letztlich nichts Neues, sondern nur eine andere Formulierung dafür, dass Korrelationsstudien eben Korrelationen ermitteln, und Korrelationen noch keinen Kausalzusammenhang belegen (vgl. oben bei Fn. 20).

Die Frage, ob die höhere Vergütung von Vorständen zu höheren Leistungen führt, betrifft gerade einen solchen Kausalzusammenhang – wie fast alle interessanten empirischen Fragen. Um sie zu beantworten, müssten neben der Korrelation auch noch die Nachzeitigkeit und das Fehlen von Verunreinigungen festgestellt werden (oben A.II.). Ersteres ist zwar durch Panelstudien möglich (dazu § 2 bei Fn. 78), weshalb Korrelationsstudien in

---

[64] Nachfolgende Darstellung orientiert an *Börsch-Supan/Köke*, Ger Econ Rev 2002, 295 (Fazit 322: “what appear to be technical problems […] are in fact deep problems of any empirical analysis, independent of the technical tool.”).

[65] Weniger trennscharf führt *Abelson*, Principled Argument 1995, 181 Fn. 7 die Kausalitätsumkehr als Sonderfall der “Third Variable” (180 ff.) ein, die im nächsten Satz des Haupttextes besprochen wird.

[66] Instruktiv *Devers/Cannella u.a.*, J Mgmt 2007, 1016, 1018 ff.

[67] Dazu *Börsch-Supan/Köke*, Ger Econ Rev 2002, 295, 297 f.; “spurious” heißt „unehelich“ (von lat. *spurius*, uneheliches Kind, *Harper*, Etymology 2014, zum „Spurius“ auch *Dudenredaktion*, Fremdwörterbuch 2000, 1264), was besser passt als die übliche Übersetzung „Scheinkorrelation“, denn die Korrelation existiert nicht scheinbar, sondern tatsächlich – und stammt nur aus anderer Quelle als vermutet.

[68] *Atteslander*, Methoden 2010, 23 ff.; *Diekmann*, Sozialforschung 2012, 68 f.; *Matthews*, Teach Stat 2000, 36, passim; *Kronmal*, J Royal Stat Soc A 1993, 379, 382 ff.; im juristischen Schrifttum *Röhl*, Rechtssoziologie 1987, 116; *Petersen*, Staat 2010, 435, 448; *Petersen/Goerg* in: Towfigh/Petersen, Methoden 2010, 201, 234 f.

[69] Zu diesem in den USA beliebten Bsp. *Lawless/Robbennolt/Ulen*, Methods 2010, 299.

diesem Format durchaus die höchste Eindeutigkeit geltend machen können[70] und dementsprechend auch in der Forschung zur Vorstandsvergütung immer öfter eingesetzt werden.[71] Die Panelforschung begegnet aber ganz eigenen Herausforderungen, insbesondere kann sie den Teilnehmern kaum verborgen bleiben (sondern findet meist im Rahmen von Befragungsstudien statt, dazu unten bei Fn. 258) und kann sie durch systematischen Stichprobenausfall (*sample attrition*, Panelmortalität) verzerrt werden (*unbalanced panel*).[72] Zudem können auch Panelstudien die Nachzeitigkeit nie ganz sicherstellen – wenn beispielsweise die Vorstände im Panel alle Änderungen der Vergütungshöhe stets vorhersähen, könnte die spätere Vergütungshöhe auf die aktuelle Leistung *vor*wirken.

Gar nicht ausschließen kann die Korrelationsstudie jedenfalls, dass andere Einflussfaktoren den Rückschluss auf die erklärende Variable verunreinigen. Die Menge der aufgezeichneten Daten über einen vergangenen Lebensvorgang, die Korrelationsstudien zugrunde liegt, ist naturgemäß beschränkt. Was sie beispielsweise fast nie enthält, sind Informationen darüber, wie die Menschen, die zum Studienzeitpunkt in der beobachteten Situation vorgefunden wurden, dorthin gekommen sind. Meistens entscheiden sie sich ja bewusst für oder gegen eine bestimmte Situation (*self selection*).[73] Falls beispielsweise kompetentere Vorstandsmitglieder gezielt Arbeitgeber suchen, deren Aufsichtsratsvorsitzender ihnen persönlich nahesteht, wäre die festgestellte Korrelation zwischen Vergütung und Leistung vielleicht sogar dann zu beobachten, wenn die Vergütung als solche keinerlei Anreizwirkung entfaltete.

Aus diesen Gründen beschränkt sich die Studie von Kaplan und Rauh zu Recht darauf, einen „starken Zusammenhang" zwischen Vergütung und Leistung festzustellen, ohne etwas über die Richtung des Ursachenzusammenhangs auszusagen.[74]

---

[70] *Börsch-Supan/Köke*, Ger Econ Rev 2002, 295, 301: "Panel data are a necessary condition for the identification of any singular corporate governance mechanism".

[71] *Devers/Cannella u.a.*, J Mgmt 2007, 1016, 1041: "compensation research is moving away from an emphasis on cross-sectional research and toward longitudinal and dynamic panel research."

[72] *Bortz/Döring*, Forschungsmethoden 2006, 447 ff. m.w.N. zur Lehrliteratur.

[73] Näher *Diekmann*, Sozialforschung 2012, 39 f., 64 f.; *Legewie*, KZfSS 2012, 123, 128.

[74] Vgl. Zitat oben Fn. 62; "pay-for-performance" deutet zwar eine Ursachenrichtung von Leistung zu Vergütung an, meint aber nur den "link" zwischen beiden Größen (*Kaplan/Rauh*, Rev Fin Stud 2010, 1004, 1044); ähnlich vorsichtig *Rapp/Schaller/Wolff*, DBW 2011, 311, 327: „verzichten wir bewusst auf eine detaillierte Interpretation des Zusammenhangs […] aktienkursbasierter Langfristanreize mit der vergangenen Performance."

b) Auswahlfehler (sample selection bias)

Die eben angesprochene Selbstauswahl ist nur ein spezieller Fall einer allgemeineren Beschränkung von Korrelationsstudien, nämlich ihrer nahezu unausweichlichen Verunreinigung durch Auswahlfehler. Gerade von den empirischen Studien zur *corporate governance* – zu denen auch diejenige von Kaplan und Rauh gehört – unterliegen „wahrscheinlich die meisten" und jedenfalls „alle aus Deutschland" solchen Fehlern.[75] Am häufigsten entstehen sie durch die Verwendung sog. Gelegenheits- oder Bequemlichkeitsstichproben (*opportunity* bzw. *convenience sample*), also der am einfachsten verfügbaren Daten.

Da viele Daten überhaupt nur von börsennotierten Unternehmen offengelegt werden müssen (nicht zuletzt aufgrund der strengen deutschen Datenschutzvorschriften)[76], sind Unternehmensdaten nicht gleichmäßig und auch nicht in gleicher Genauigkeit zugänglich. Selbst unter den börsennotierten Unternehmen gibt es noch deutliche Abstufungen,[77] daher untersuchen Korrelationsstudien fast immer – Kaplan und Rauh machen keine Ausnahme – nur die größten oder erfolgreichsten börsennotierten Unternehmen, die in einem Auswahlindex zusammengefasst und deren Daten zuverlässig publiziert sind. Gleichzeitig sind Größe und Erfolg aber immer entweder die zu erklärende Variable oder mit dieser Variable untrennbar verbunden – so auch im Fall der Vorstandsvergütung.[78] Daher stellen Unternehmensgröße und -erfolg (sowie eine Vielzahl damit verbundener Einflussgrößen) stets Verunreinigungen dar, die Aussagen über Ursachenbeziehungen grundsätzlich vereiteln.

Zur konkreten Stichprobe von Kaplan und Rauh finden sich in einer späteren Studie Belege für systematische Unterschiede zwischen Unternehmen aus der ExecuComp-Datenbank und anderen,[79] die den Ursachenschluss verunreinigen könnten. Da hier aber die Kritik an der konkreten

---

[75] *Börsch-Supan/Köke*, Ger Econ Rev 2002, 295, 305: "most studies of corporate governance are likely to have a sample selection bias." und weiter 306: "To our knowledge, all empirical studies on corporate governance in Germany suffer from the econometric problem of selection bias."

[76] Darauf spielen auch *Börsch-Supan/Köke*, Ger Econ Rev 2002, 295, 307 an.

[77] Vgl. *Reiner*, AG 2006, 93, 103: Die kostenlose Bereitstellung seiner Satzung unterließ „kein einziges DAX-Unternehmen, eine mäßige Zahl [8 %] der im MDAX und eine erschreckend hohe Zahl [24 %] der im SDAX vertretenen Gesellschaften".

[78] *Börsch-Supan/Köke*, Ger Econ Rev 2002, 295, 305: "data on management compensation are available only for those firms which enacted compensation programs. But these enterprises tend to be well-performing firms, often listed on the stock exchange."

[79] *Cadman/Klasa/Matsunaga*, Acc Rev 2010, 1511, 1542: "differences in contracting environments between ExecuComp and non-ExecuComp firms lead to differences in the structure of CEO incentive contracts […] Researchers should use caution when generalizing results to firms that are not included in the ExecuComp data set."

Studie weniger interessiert als die grundsätzliche methodische Beschränkung, genügt es festzuhalten, dass Erkenntnisse aus Korrelationsstudien grundsätzlich durch Auswahlfehler verunreinigt sein können, weil sie nur auf verfügbare Daten zurückgreifen können.[80] Zugleich beschränken Auswahlfehler die Übertragbarkeit der Erkenntnisse auf Unternehmen außerhalb der Bequemlichkeitsstichprobe,[81] wenngleich eine auf „DAX-Unternehmen zum aktuellen Zeitpunkt" beschränkte Aussage natürlich auch ihren Wert haben mag.

c) Spezifikationsfehler (missing variables)

Empirische Hypothesen müssen spezifizieren, wie sich die abhängige Variable ändert, wenn sich die unabhängigen Variablen ändern (oben § 2 bei Fn. 87). Werden dabei entscheidende Variablen übersehen, liegt ein Spezifikationsfehler vor. Das anschaulichste Beispiel für dieses Problem ist das Yule-Simpson-Paradox,[82] das herkömmlich als „Simpson-Paradox" bekannt ist, wenngleich es nach seiner Entwicklungsgeschichte „Pearson–Yule–Kendall–Simpson-Paradox" heißen sollte.[83] Seine berühmteste Demonstration stammt aus der Studienzulassungspraxis der Universität Berkeley,[84] es taucht aber auch in anderen Zusammenhängen auf[85] und kann sogar einen vermeintlich gesundheitsförderlichen Effekt des Rauchens belegen.[86] Vorliegend genügt ein vereinfachtes Beispiel:

Stellen wir uns eine dichotome Welt vor, in der Vorstände nur entweder hoch oder niedrig bezahlt sein können und nur entweder Leistung bringen oder faulenzen (*shirking*) können. Nun beobachten wir 2.200 Vorstände, von denen die Hälfte hoch bezahlt und die Hälfte niedrig bezahlt ist. Von den 1.100 hoch Bezahlten bringen 610 Leistung, während 490 faulenzen, von den 1.100 niedrig Bezahlten bringen 390 Leistung, während 710 faulenzen. Alles deutet darauf hin, dass hohe Bezahlung leistungsförderlich ist: Nur 35 % der niedrig Bezahlten, aber ganze 55 % der hoch Bezahlten bringen Leistung. Bis jemand feststellt, dass es in unserer dichotomen Welt auch zwei verschiedene Branchen gibt: Automobil und Einzelhandel.

---

[80] *Börsch-Supan/Köke*, Ger Econ Rev 2002, 295, 311: "we usually use data that were collected for other purposes and hence are incomplete for our means."

[81] Vgl. schon oben § 2 bei Fn. 263 sowie *Wheeler* in: Cane/Kritzer, Hdb Empirical 2010, 125, 147: "it seems that we know almost nothing about what occurs in firms that are not listed on the stock market".

[82] Dazu *Krämer*, Denkste! 2011, 168 ff. mit Grafik auf 179.

[83] *Good/Mittal*, Ann Stat 1987, 694, 695.

[84] *Bickel/Hammel/O'Connell*, Sci 1975, 398.

[85] Übersicht bei *Ooi*, Wharton Res Sch J 2004.

[86] *Dubben/Beck-Bornholdt*, Fehlinformation 2011, 232 m.Verw. auf *Appleton/French/Vanderpump*, Am Stat'n 1996, 340.

Nun betrachten wir diese beiden Branchen näher, und stellen fest: Von den 100 niedrig bezahlten Einzelhandelsvorständen faulenzen nur 10 %, von den 1.000 hoch Bezahlten dagegen 40 %. Ähnliches beobachten wir in der Automobilbranche: von ihren 100 hoch bezahlten Vorständen faulenzen sogar ganze 90 %, von den 1.000 niedrig Bezahlten aber nur 70 %. Obwohl also insgesamt ein geringerer Anteil der hoch Bezahlten faulenzt, ist die Faulenzerquote der hoch Bezahlten in *beiden* Branchen höher als die der niedrig Bezahlten! Der vermeintlich positive Vergütungseffekt im Aggregat resultiert einzig und allein aus dem unterschiedlichen Grundanteil der hoch Bezahlten (über 90 % im Einzelhandel, unter 10 % in der Automobilbranche).

	Bezahlung				
Leistung	niedrig			hoch	
N – niedrig	710	65 %	>	45 %	490
H – hoch	390	35 %	<	55 %	610
		=			=
$N_{\text{Nur Automobil}}$	700	70 %	<	90 %	90
$H_{\text{Nur Automobil}}$	300	30 %	>	10 %	10
		+			+
$N_{\text{Nur Einzelhandel}}$	10	10 %	<	40 %	400
$H_{\text{Nur Einzelhandel}}$	90	90 %	>	60 %	600

Illustration: Die Addition von Teilwerten (Kleindruck außen) kann prozentuale Trends (innen) völlig umkehren. (Prozentwerte geben hier das Verhältnis hoch und niedrig Bezahlter in den sechs Kombinationen aus Bezahlung und Branche an.)

Das Yule-Simpson-Paradox beschreibt eine eigenwillige, aber keineswegs rein theoretische Situation.[87] In Konstellationen mit zwei Einflussfaktoren, wie der eben skizzierten, ist schon nach dem Zufall in jeder sechzigsten Situation mit einer solchen Umkehrung zu rechnen.[88] Die negativen Konsequenzen eines Spezifikationsfehlers beschränken sich aber nicht auf den Extremfall der Umkehrung. Vielmehr fehlt die dritte Voraussetzung für Ursachenschlüsse (oben bei Fn. 17) generell, wenn der Einfluss unbeobachteter Störvariablen nicht ausgeschlossen werden kann.[89] Anders gewendet: Das Yule-Simpson-Paradox entsteht, weil und wenn die Verteilung der vermeintlich erklärenden Variable (hier Vergütungshöhe) ihrerseits von einer Störvariable bestimmt wird. Diese Störvariable war im obigen Beispiel die Branchenzugehörigkeit, die in der Realität wohl niemand übersehen dürfte, aber die gleiche Wirkung kann von jeder anderen Einflussgröße ausgehen, die in der formulierten Hypothese fehlt.[90] Die in Korrelationsstudien beobachteten Lebenssachverhalte sind viel zu komplex, um alle mög-

[87] Vgl. *Wagner*, Am Stat'n 1982, 46, 47: "Simpson's paradox is not a contrived pedagogical example. Because this situation occurs at the level of a purely descriptive data analysis, it can easily bewilder the statistically naive observer."

[88] *Pavlides/Perlman*, Am Stat'n 2009, 226, 228: "prior probability of Simpson's Paradox in a 2×2×2 table under the uniform prior is 0.0166 (correct to three significant figures)."

[89] *Börsch-Supan/Köke*, Ger Econ Rev 2002, 295, 313; zum Begriff der Störvariable oben § 2 bei Fn. 69.

lichen Einflussgrößen auch nur benennen, geschweige denn quantifizieren zu können.[91] Deshalb wird bisweilen sogar bezweifelt, dass reine Beobachtungsdaten mit den Methoden der Statistik überhaupt sinnvoll auswertbar sind.[92]

Spezifikationsfehler gehören zu den größten Herausforderungen der beobachtenden Forschung. Da schon denklogisch kein statistischer Test ermitteln kann, welche Variablen übersehen wurden, lassen sich Spezifikationsfehler überhaupt nur durch die Berücksichtigung theoretischen Vorwissens minimieren.[93] Damit wächst für jede beobachtende Studie die Begründungslast, weil sie im Prinzip alle in früheren Studien erkannten Einflussgrößen miterheben und berücksichtigen muss.[94] Für Kapitalmarktstudien (wie jene von Kaplan und Rauh) bedeutet das, dass zumindest Daten über die Anteilseignertypen und Größe von Anteilsblöcken, über verschiedene Aspekte der Finanzierungsstruktur und die Zusammensetzung der Leitungsorgane, die Marktmacht und Wettbewerbsstellung sowie die Anzahl der erfolgreichen und erfolglosen Übernahmeversuche berücksichtigt werden sollten.[95] Es sei hier dahingestellt, welche dieser Einflussgrößen die vorliegende Studie tatsächlich erhebt – es sind nicht viele – und welche weiteren für eine belastbare Aussage über den Zusammenhang zwischen Vorstandsvergütung und Leistung erforderlich gewesen wären.

d) Messfehler (measurement error)

Die vierte methodische Beschränkung von Korrelationsstudien liegt in der unvermeidbaren Ungenauigkeit ihrer Messinstrumente. Das Problem besteht allgemein insoweit als eine mangelhafte Operationalisierung immer mangelhafte Daten ergeben wird, verstärkt sich aber in Korrelationsstudien, weil nur eine beschränkte Menge an Daten verfügbar ist.

Im Unternehmenskontext ist insbesondere unklar, wie die Zielgröße (hier: Vorstandsleistung) überhaupt operationalisiert werden kann. Da die

---

[90] Bsp. bei *Börsch-Supan/Köke*, Ger Econ Rev 2002, 295, 312: “An interesting example of a frequently missing variable is the intensity of product market competition” m.w.N.

[91] *Kanazawa*, Ration Soc 1999, 443, 453 ff. bezieht das sogar auf Experimente: “what if the treatment effects interact with some *unknown* subject characteristics?”

[92] *Salsburg*, Statistics 2001, 302 f.; *Gorard*, Brit J Sociol Edu 2006, 67, 71: “If the researcher does not use a random sample then traditional statistics are of no use since the probabilities then become meaningless.”

[93] *Börsch-Supan/Köke*, Ger Econ Rev 2002, 295, 315: “Typically, the researcher cannot test for missing variables. […] we must rely on economic theory to detect which variables are missing”.

[94] Das erklärt die andere Stoßrichtung, die dem Yule-Simpson-Paradox oft gegeben wird: Wann dürfen mehrere unterschiedliche Datenmengen vereinigt (*pooled*) werden? Dazu schon *Bickel/Hammel/O’Connell*, Sci 1975, 398, 401 f.

[95] So die Aufzählung bei *Börsch-Supan/Köke*, Ger Econ Rev 2002, 295, 315.

äußere Gültigkeit im UTOS-Schema (oben bei Fn. 22) auch von der Übertragbarkeit auf der Zielgrößen-Dimension (O) abhängt, sollten empirische Erkenntnisse nicht davon abhängen, wie die Zielgröße (hier Vorstandsleistung) gemessen wird. Will man beispielsweise – wie es häufig geschieht – die Vorstandsleistung am Unternehmenserfolg messen, muss man sich für eine der denkbaren Erfolgskennzahlen entscheiden, wie beispielsweise das Kurs-Gewinn-Verhältnis (KGV, engl. *price earnings ratio* PER), die Gesamtkapitalrentabilität (GKR, engl. *return on assets* ROA), die Eigenkapitalrentabilität (EKR, engl. *return on equity* ROE) oder das Kurs-Substanzwert-Verhältnis (Tobin's q). Allerdings korrelieren diese Kennzahlen untereinander kaum,[96] also kann von einer konsistenten Messung des Unternehmenserfolgs kaum die Rede sein[97] – ähnlich wie aus der Praxis der Unternehmensbewertung bekannt ist, dass verschiedene Bewertungsmethoden zu ganz unterschiedlichen Wertschätzungen führen:[98] „den einen ‚richtigen' Unternehmenswert gibt es deshalb nicht."[99]

Nichts anderes gilt letztlich für die Operationalisierung des Begriffs „Vorstandsleistung" bei Kaplan und Rauh. Die Autoren behaupten zwar nicht, dass die von ihnen betrachtete Kapitalrendite unmittelbar die Leistung des Vorstands abbildet; will man ihre Studie aber für die vorliegende Frage fruchtbar machen, ist zu reflektieren, dass jedes Maß des Unternehmenserfolgs von zahlreichen Faktoren abhängt, die nicht in der Hand des Vorstands liegen.[100] Ebenso problematisch wie die Operationalisierung der abhängigen Größe (Vorstandsleistung) ist die Operationalisierung der unabhängigen Größe (Vorstandsvergütung); insbesondere ist unklar, welche Vergütungsbestandteile berücksichtigt werden müssen und wie sie quantifiziert werden können.

Schließlich verbleibt selbst bei der besten Operationalisierung immer noch das statistische Grundrauschen einer unbekannten Vielzahl moderie-

---

[96] *Börsch-Supan/Köke*, Ger Econ Rev 2002, 295, 318 f. mit dem Fazit: "different measures of performance might be correlated only in a highly non-linear way."; vgl. auch noch § 4 bei Fn. 332.

[97] So auch der Forschungsbericht *Devers/Cannella u.a.*, J Mgmt 2007, 1016, 1039: "we found little consistency in the operationalization of many important constructs of interest."

[98] Nach *Stilz* in: Spindler/Stilz, AktG 2010, § 255 Rn. 20 m.w.N. kommen „auch bei methodisch sauberem und vergleichbarem Vorgehen aufgrund der in zahlreichen Punkten erforderlichen Prognoseentscheidung mehrere Gutachten kaum je zum selben Ergebnis"; ausf. *Veil* in: Spindler/Stilz, AktG 2010, § 305 Rn. 44 f.; *Wangler*, Sachvst 2005, 245.

[99] *Stephan* in: Lutter/Schmidt, AktG 2010, § 305 Rn. 61 m.w.N. aus der Rspr. in Fn. 106.

[100] *Devers/Cannella u.a.*, J Mgmt 2007, 1016, 1021 m.w.N.: "considerable literature suggests that firm performance is not only a function of managerial decisions, but also of factors outside managers' control", deshalb sei die "pay-to-performance relationship" sehr "coarse and distal" (1025).

render Einflüsse, das den Ursachenschluss erschwert.[101] Dieses Störrauschen lässt sich unter Umständen reduzieren, aber nie ganz beseitigen.

III. Laborexperimente – Idealtyp der Eindeutigkeit

Das größtmögliche Maß an Eindeutigkeit können Studien erzielen, die in einer genau kontrollierten Umgebung Daten erheben. Am weitesten geht diese *Kontrolle* – bezogen sowohl auf die Kontextfaktoren der Datenentstehung als auch auf die Art und Genauigkeit der Datenaufzeichnung – in Laborexperimenten. Indem sie es ermöglichen, die zeitliche Reihenfolge von Ereignissen zu manipulieren und Moderatorvariablen (dazu oben § 2 bei Fn. 69) entweder auszuschließen oder mitzuerheben, können Laborexperimente die drei Voraussetzungen der inneren Gültigkeit (oben A.I.) am besten erfüllen und dadurch Erkenntnisse mit einem Höchstmaß an Eindeutigkeit gewinnen.[102]

1. Ein Beispiel: Ariely, Gneezy, Loewenstein & Mazar (2009)

Als gutes Beispiel für Laborexperimente mag die 2009 veröffentlichte Studie zweier Psychologen gemeinsam mit einem Volkswirt und einer Betriebswirtin dienen.[103] Dass Unternehmensleiter ebenso wie viele andere Berufsgruppen erfolgsbasiert entlohnt werden, um sie zu höherer Leistung anzuhalten,[104] beruht nach dieser Studie auf zwei Annahmen: Erstens dass höhere Bezahlung stärker motiviere, und zweitens dass höhere Motivation auch zu höherer Leistung führe; beides folge zwanglos aus dem ökonomischen Rationalmodell, sei aber empirisch nicht durchweg unzweifelhaft.[105] Nachdem drei der vier Autoren in anderen Studien schon die erste der beiden Annahmen in Zweifel gezogen hatten, hinterfragten sie in der vorliegenden Studie nun die zweite Annahme. Denn angesichts früherer Studien zu ähnlichen Themen war durchaus fraglich, ob finanziell stimulierte Moti-

---

[101] *Börsch-Supan/Köke*, Ger Econ Rev 2002, 295: “estimation results are likely to be weakened owing to a low ratio of signal-to-noise.”

[102] *Shadish/Cook/Campbell*, Causal Inference 2002, 7: “No other scientific method regularly matches the characteristics of causal relationships so well.”; vgl. auch *Angrist/Pischke*, Econometrics 2009, 11: “The most credible and influential research designs use random assignment.” und weiter 15: “in principle they solve the most important problem that arises in empirical research.”

[103] *Ariely/Gneezy u.a.*, Rev Econ Stud 2009, 451; Forschungsbericht über Experimente zur Frage “What sort of management compensation contract is most likely to align managerial actions with shareholder preferences?” in *Cadsby/Maynes*, Manage Dec Econ 1998, 277, 280 ff.

[104] *Ariely/Gneezy u.a.*, Rev Econ Stud 2009, 451: “Many, if not most upper-management […] and workers in a wide variety of other jobs are rewarded for their effort […] to motivate individuals to increase their effort, and hence their output”.

[105] *Ariely/Gneezy u.a.*, Rev Econ Stud 2009, 451 f. m.w.N.

vation tatsächlich größere Leistungen bewirkt: Schon die 1908 mit Mäusen durchgeführten Experimente von Yerkes und Dodson deuteten darauf hin, dass *zu starke* physische Anreize die Bewältigung schwieriger Aufgaben vielleicht eher behindern als erleichtern.[106] Spätere Experimente in der Sozialpsychologie belegten fünfundzwanzig Jahre vor Ariely und Kollegen ein ähnliches „Versagen unter Erfolgsdruck“ (*choking under pressure*),[107] das vor allem in der Sportforschung Beachtung fand.[108]

Ob auch hohe finanzielle Anreize solches Versagen begünstigen, oder ob sie die Leistung eher steigern, untersuchten die Autoren in drei verschiedenen Experimenten, deren Teilnehmer unterschiedliche Aufgaben durch physische oder kognitive Anstrengung bewältigen mussten. Zwei der Experimente variierten die Höhe der zu erwartenden Vergütung, stellten ihren Teilnehmern also einmal einen geringen Lohn in Aussicht und einmal einen (zehn- oder hundert Mal so) hohen. Das erste dieser beiden Experimente fand in einer ländlichen Gegend Indiens statt, deren Preisniveau erhebliche finanzielle Anreize ermöglichte: Die teilnehmenden Dorfbewohner konnten mit weniger als einer Stunde Zeiteinsatz einen Geldbetrag verdienen, der den ortsüblichen Pro-Kopf-Verbrauch entweder für anderthalb Tage, einen halben Monat oder für fast fünf Monate deckte (Querschnitt, *between subjects*).[109] Für das zweite Experiment wurden Studenten der Technischen Hochschule Massachusetts (MIT) rekrutiert,[110] die in etwa einer Viertelstunde bis zu 48 Euro und in einer weiteren Viertelstunde bis zu 480 Euro verdienen konnten (Längsschnitt, *within subject*).[111]

Zumindest für kognitiv anspruchsvolle Aufgaben stellten die Autoren in beiden Experimenten fest, dass die Versuchsteilnehmer bei einer auf das zehn- bzw. hundertfache *erhöhten* Höchstvergütung einen durchschnittlich

---

[106] Zit. bei *Ariely/Gneezy u.a.*, Rev Econ Stud 2009, 451, 452; online verfügbar bei *Green*, Classics 2012.

[107] *Ariely/Gneezy u.a.*, Rev Econ Stud 2009, 451, 452 m.Verw. auf *Baumeister*, J Pers Soc Psy 1984, 610 (dort insb. 616 f., Experiment 5); jünger *Beilock/Carr*, Psy Sci 2005, 101; vgl. auch *Markman/Maddox/Worthy*, Psy Sci 2006, 944 zum spiegelbildlichen “excelling under pressure”.

[108] Z.B. *Beilock/Carr*, J Exp Psy Gen 2001, 701 (Golf); *Dohmen*, J Econ Behav Org 2008, 636 (Fußball); *Hill/Hanton u.a.*, Int Rev Sport Exerc Psy 2010, 24 (Forschungsbericht).

[109] Laut *Ariely/Gneezy u.a.*, Rev Econ Stud 2009, 451, 454 f. war der Höchstverdienst “approximately equal to half of the mean yearly consumer expenditure” – die berichteten Werte (6 x 400 / 495) ergeben 4,85 Monate. Allgemein zur Querschnittstudie oben § 2 bei Fn. 76.

[110] *Ariely/Gneezy u.a.*, Rev Econ Stud 2009, 451, 461: “The experiment was conducted towards the end of the semester, a time when the students are likely to have depleted their budget and be strapped for cash.”

[111] Das sind für zwei verschiedene Aufgaben je einmal 30 und einmal 300 US$, umgerechnet zum durchschnittlichen Wechselkurs (0,804) der Jahre 2003–2007, in denen das Experiment vermutlich stattfand. Allgemein zur Längsschnittstudie oben § 2 bei Fn. 76.

14 bis 24 Prozentpunkte *geringeren* Anteil dieser Vergütung erzielten.[112] Die höhere in Aussicht gestellte Vergütung führte also paradoxerweise zu einem geringeren Maß an Leistung, daher schlussfolgerten die Autoren,

> „dass nicht angenommen werden darf, dass Anreize zu setzen oder zu erhöhen immer die Leistung fördert. Jenseits einer gewissen Schwelle, so scheint es, können höhere Anreize die Motivation auf ein über-optimales Niveau steigern und perverse Auswirkungen auf die Leistung zeitigen.“[113]

2. Methodische Beschränkungen

Die Studie von Ariely und Kollegen ist in mancher Hinsicht typisch, in anderer aber ungewöhnlich für die experimentelle Herangehensweise, und eignet sich daher gut als Prüfstein dieser Methode. Erneut werde ich nicht näher auf mögliche Kritikpunkte an der konkreten Studie eingehen,[114] sondern vielmehr die Möglichkeiten und Beschränkungen der Methode allgemein erläutern.

Wie bei der Korrelationsstudie, so sind auch hier die methodischen Beschränkungen am besten als Kehrseite der Möglichkeiten zu verstehen. Das Laborexperiment ermöglicht gezielte Manipulation und damit maximale Kontrolle über einen Lebenssachverhalt. Umgekehrt mag diese Kontrolliertheit so künstlich und unrealistisch erscheinen, dass sie Zweifel an der Übertragbarkeit der so gewonnen Erkenntnisse weckt.[115] Das betrifft im

---

[112] Eigene Auswertung der Daten in *Ariely/Gneezy u.a.*, Rev Econ Stud 2009, 451, 457 Tabelle 3 (Z. 1–3, Sp. 1 & 3) und 462 Haupttext: kleinster Unterschied 53,6 - 39,7 = 13,9, größter Unterschied 48,2 - 24,1 = 24,1.

[113] *Ariely/Gneezy u.a.*, Rev Econ Stud 2009, 451, 467.

[114] Z.B. bleibt darin völlig offen (1) warum in rein physischen Aufgaben die MIT-Studenten mit steigendem Anreiz mehr leisteten, die Inder dagegen weniger, und (2) warum in kognitiven Aufgaben die Leistung der Studenten schon bei zehnfachem Anreiz sank, die der Inder dagegen erst bei hundertfachem (also nur bei der zweiten Verzehnfachung).

[115] Das gilt (nur) als unbedenklich, soweit ein Experiment allgemeine Verhaltenstheorien (oben § 2 A.VII.3.) testet, vgl. *Mook*, Am Psy’st 1983, 379, 380: “We are not *making* generalizations, but *testing* them.” und weiter 383: “we are not using what happens in the laboratory to ‘predict’ the real world. Prediction goes the other way: Our theory specifies what subjects should do in the laboratory. Then we go to the laboratory to ask, Do they do it? And we modify our theory, or hang onto it for the time being, as results dictate.”; zust. *Thye* in: Webster/Sell, Experiments 2007, 83; *Zeiler*, JITE 2010, 178, 182 f.; a.A. *Hogarth*, J Acc Res 1982, S108, 111: “artificial experiments carry a particularly strong bias toward reinforcing the status quo of current knowledge.”; ausf. *Zelditch* in: Webster/Sell, Experiments 2007, 87 ff. (Kap. 4).

Prinzip jede der vier UTOS-Dimensionen,[116] doch besondere Skepsis gilt häufig der Situations-Dimension (S) und der Teilnehmer-Dimension (U):

> „Das Problem der Übertragbarkeit von Laborbefunden ist also ein Doppeltes: Zum einen geht es darum, ob die Laborbedingungen die Realität außerhalb des Labors widerspiegeln. […] Zum anderen stellt sich die Frage, ob sich das Verhalten der Probanden verallgemeinern lässt."[117]

Jede dieser beiden Beschränkungen bespreche ich nun allgemein sowie mit Blick auf ihre jeweils wichtigste Erscheinungsform.

a) Abstraktion vom Lebenssachverhalt

Die erste Beschränkung ist zugleich die meistzitierte: Die Versuchsteilnehmer experimenteller Studien befinden sich in *künstlichen Situationen*, die oberflächlich betrachtet meist überhaupt nichts mit der Situation zu tun haben, aus der die jeweilige Forschungsfrage stammt.[118] Dafür ist die Studie von Ariely und Kollegen typisch. Nehmen wir nur eine der Aufgaben, mit denen die Versuchsteilnehmer des zweiten Experiments konfrontiert wurden: Sie mussten aus einer 3x4-Tabelle mit zwölf rationalen Zahlen schnellstmöglich die beiden Zahlen heraussuchen, die in der Summe zehn ergeben.[119] Das tun Vorstände nie. Es „kommt einer einfachen logischen Rechenaufgabe gleich. Unternehmerisches Handeln weist indes eine deutlich höhere Komplexität auf",[120] daher könnte man meinen, die Versuchssituation sei vom „echten" Leben kategorisch verschieden (ein *aliud*) und auf dieses daher gar nicht übertragbar. Andererseits bildet die Versuchssituation von Ariely und Kollegen eine Klasse von Herausforderungen ab, die auch Vorstände regelmäßig meistern müssen: Das Lösen von Problemen. „Alles Leben ist Problemlösen",[121] und die Vorstandssituation hat mit der Versuchssituation wenigstens gemein, dass sie kognitiv anspruchsvolle Problemlösungen erfordert – und womöglich gehen die Parallelen im Einzelfall sogar noch viel weiter.[122] So betrachtet erscheint die Versuchssituation vom „echten" Leben nur graduell verschieden (als *minus*). Man könnte

---

[116] Dazu oben bei Fn. 22; vgl. *Dipboye/Flanagan*, Am Psy'st 1979, 141, 146: "The question of external validity must begin with a delineation of the actors, settings, and behaviors to which one wishes to generalize"; ähnl. *Hussy/Schreier/Echterhoff*, Forschungsmethoden 2010, 131 (Populations-, Situations- und Variablenvalidität).

[117] *Englerth*, Verbrecher 2010, 376; ebenso schon *Winer*, J Acad Mktg Sci 1999, 349, 351; *Croson*, U Ill L Rev 2002, 921, 942 f.

[118] Krit. *Stanovich*, Psychology 2012, 108: gleiches gelte etwa für die Teilchenphysik.

[119] Bsp.: „9,38 – 6,74 – 8,17 || 5,15 – 6,61 – 3,06 || 9,71 – 0,91 – 4,88 || 3,58 – 4,87 – 6,42" (*Ariely/Gneezy u.a.*, Rev Econ Stud 2009, 451, 460).

[120] So *Englerth* in: Engel u.a., Recht und Verhalten 2007, 60, 74 zu einem ähnlichen Experiment; vgl. noch § 4 bei Fn. 615.

[121] So der tiefgründige Titel der Werkauswahl von *Popper*, Problemlösen 2001.

sagen, die experimentelle Aufgabe abstrahiere von unnötigen Details der Vorstandsarbeit. Das wirft zwei Fragen auf: Warum abstrahieren? Was ist unnötig?

Jede empirische Erkenntnis beruht auf Abstraktion. Ganz gleich wie komplex das Gehirn und wie umfangreich seine Erkenntnisfähigkeit, findet es sich doch *innerhalb* einer Umwelt, die demnach komplexer sein muss als das Gehirn und von ihm nie vollständig begriffen werden kann.[123] Schon jede Bezeichnung oder Beschreibung der Realität mittels der Sprache reduziert lebensweltliche Komplexität.[124] Zu Recht wird daher festgestellt: „Die Komplexität der Lebenswirklichkeit ist […] für alle Sozialwissenschaftler ein epistemologisches Problem.“[125] Das gilt indessen nicht nur für Sozialwissenschaftler, sondern für alle Realwissenschaften. Auch die Physik als wissenschaftliche Königsdisziplin kann das Verhalten physikalischer Systeme auf der Makroebene nie vollständig vorhersagen[126] und ist auf Abstraktionen angewiesen. Ein berühmter Physikerwitz definiert den Begriff „Kuh“ als „eine Kugel, die in alle Richtungen isotrop Milch abgibt… zumindest in erster Näherung.“ Dieser Abstraktionsschritt gilt als „vielleicht das Schwierigste und zugleich Wichtigste, wenn man Physik lernen will“:[127]

„ganz wichtig ist für das tagtägliche Denken in der Wissenschaft: *Bevor du irgend etwas tust, abstrahiere von allen irrelevanten Details!* […] Die Befreiung von irrelevanten Details ist der erste Schritt, wenn man irgendein Weltmodell bauen will, und das tun wir unbewußt schon von dem Moment an, wenn wir geboren werden. Es aber bewußt zu tun, ist eine ganz andere Sache. [… Dabei] liegt wahrscheinlich die größte Herausforderung in der Wahl, auf welchem Weg man an ein Problem herangehen soll. Die bloße Beschreibung einer Bewegung entlang einer Geraden – eines der grundlegenden Prinzipien in der

---

[122] Leicht vorstellbar sind Situationen, in denen die Leistung des Vorstands davon abhängt, Bilanzzahlen unter Zeitdruck zu beurteilen – *whether they "add up"*, wie es im Englischen seit 1942 bezeichnenderweise heißt (*Harper*, Etymology 2014, Stichwort "add").

[123] Das Argument stammt nicht von mir, allerdings ist mir seine Originalquelle entfallen. Vgl. jedenfalls *Springer*, SdW 2012, 9/48, 49: „Die Realität ist unendlich komplex.“

[124] *Starck*, JZ 1972, 609, 610: „Die Deutung der Fakten vollzieht sich ganz grob schon in der Sprache“, und weiter 613: „Wir sagen [.] nach *Hegel* mit Hilfe der Sprache nur das Allgemeine von den Dingen, indem wir sie auf den Begriff bringen. Damit erfassen wir aber nicht die konkrete unendliche Wirklichkeit.“

[125] *Lüdemann* in: Engel u.a., Recht und Verhalten 2007, 7, 18.

[126] *Struck*, JuS 1993, 992, 994 ff.; populärwiss. ausf. *Gleick*, Chaos 1994, passim.

[127] *Krauss*, Kugel 1998, 14 f.; vgl. auch *Stanovich*, Psychology 1998, 111: "All sciences assume that certain factors are so tangential that they will make no difference in the final results. Biologists do not generally worry that the slightly different thickness of petri dishes will significantly affect the bacteria in them." und weiter 113: "the frequent complaint about the artificiality of psychological research arises from a basic misunderstanding not only of psychology, but of basic principles that govern all sciences."

gesamten Physik – erforderte so viel Abstraktion, daß es für lange Zeit auch den einsichtsreichsten Geistern verborgen blieb, bis Galilei kam."

Abstraktion ist also in der Gewinnung neuer Erkenntnis prinzipiell unvermeidlich, denn „extern valide ist nur die Realität",[128] und wissenschaftliche Forschung ohne Abstraktion wäre just so nützlich wie die Landkarte im Maßstab 1:1, die Lewis Carroll einst persiflierte.[129] Wissenschaftliche Forschung erfordert deshalb durchweg und ganz generell eine Beschränkung des Blicks.[130] Genau wie Juristen die Welt nicht im Ganzen erfassen können und sie stattdessen durch Subsumtion unter abstrakte Normen vereinfachen,[131] können auch empirische Experimente die Lebenswelt nicht „im Kleinen" nachbauen.[132] Darin liegt nicht nur keine Schwäche, sondern geradezu eine Stärke der experimentellen Methodik.[133] Obwohl beispielsweise die Lebenswissenschaften „darauf angewiesen" sind, „Teile von Organismen außerhalb ihres Funktionszusammenhangs zu beobachten", haben sie gleichwohl „zu detaillierten Kenntnissen über die Funktionsweisen des Organismus geführt".[134]

Daraus folgt, dass Experimente nicht bereits durch die Künstlichkeit ihrer Versuchssituationen diskreditiert werden, und dass anstelle einer allgemein gehaltenen Fundamentalkritik an der experimentellen Abstraktion vielmehr eine Erörterung ihrer Angemessenheit im Einzelfall stehen muss,[135] also eine Auseinandersetzung mit der Frage, *welche* Merkmale ei-

---

[128] *Christoph Engel*, in einem persönlichen Gespräch in Bonn am 28.10.2011; pointiert *Rosenblueth/Wiener*, Phil. Sci. 1945, 316: "the best material model for a cat is another, or preferably the same cat".

[129] *Carroll*, Sylvie & Bruno Concluded 1893, 169: "'[…] And then came the grandest idea of all! We actually made a map of the country, on the scale of *a mile to the mile*!' – 'Have you used it much?' I enquired. 'It has never been spread out, yet,' said Mein Herr: 'the farmers objected: they said it would cover the whole country, and shut out the sunlight! So we now use the country itself, as its own map, and I assure you it does nearly as well. […]'"; selbe Metapher bei *Lüdemann* in: Engel u.a., Recht und Verhalten 2007, 7, 19; *Englerth*, Verbrecher 2010, 367.

[130] Dazu Zitate und Nachweise oben § 2 bei und in Fn. 244 sowie *Diekmann*, Sozialforschung 2012, 145 m.w.N.: „Theorien sollen nicht die ‚Wirklichkeit verdoppeln' […] ‚Gute Theorien sind (…) wie gute Karikaturen. Sie verzerren die Realität zwar, aber in einer Weise, dass das Wesentliche umso klarer hervortritt'."

[131] *Starck*, JZ 1972, 609, 613; *Lüdemann* in: Engel u.a., Recht und Verhalten 2007, 7, 19; zur Komplexitätsreduktion durch Rechtsnormen ausf. *von Rohr*, recht 2004, 22, passim.

[132] Missverständlich daher *Häcker* in: Bender, Tatsachenforschung 1972, 143, 151: „Im Experiment wird meist eine Miniatursituation der echten Verhältnisse geschaffen."

[133] *Stanovich*, Psychology 2012, 120: "Artificial conditions are not a drawback of experimental research. They are *deliberately* created so that we can pry variables apart."

[134] *Häcker* in: Bender, Tatsachenforschung 1972, 143, 149.

[135] *Häcker* in: Bender, Tatsachenforschung 1972, 143, 153: „Eine realistische Methodenkritik muss jedoch immer von der Frage ausgehen, ob es bessere Methoden als das Experi-

ner gegebenen Situation für den Ursachenzusammenhang relevant erscheinen, und welche nicht.[136] Dazu heißt es im bereits zitierten Text:

> „Was in einer bestimmten Situation irrelevant scheinen mag, ist keine allgemeingültige Entscheidung, sie hängt meistens davon ab, für welchen Gesichtspunkt man sich gerade interessiert."[137]

Um also zu beurteilen, inwieweit eine experimentell gewonnene Erkenntnis auf andere Kontexte übertragbar ist,[138] muss der juristische Rezipient angeben, welche Gesichtspunkte normativ „interessant" bzw. entscheidend sind.[139] Zur Studie von Ariely und Kollegen wäre also zu fragen, ob das Problemlöseverhalten der interessante Gesichtspunkt ist, wenn die Forschungsfrage lautet, ob höhere Erfolgsvergütungen die Vorstandsleistung steigern. Betrachtet man die Frage etwa im Kontext von § 87 I 1 AktG (oben bei Fn. 36), kommt es dafür entscheidend auf die Auslegung des Begriffs „Leistung" an. Die konzeptionelle Klärung dieses Begriffs durch die Rechtsdogmatik (vgl. schon oben § 2 Fn. 1) ist also Vorbedingung dafür, die Übertragbarkeit experimenteller Studien beurteilen zu können.

Alles bisher Gesagte gilt für beobachtende Studien freilich nicht minder: Auch sie abstrahieren von der Lebenswirklichkeit, und dass sie weniger abstrahieren, heißt nicht, dass sie nicht von *wichtigeren* Faktoren abstrahieren. Andererseits ist die *augenscheinliche Ähnlichkeit* ein erstes Kriterium zur Beurteilung der Übertragbarkeit (vgl. oben nach Fn. 26), daher dürfte die Argumentationslast für die Übertragbarkeit einer Studie in dem Maß ansteigen, in dem die augenscheinliche Ähnlichkeit mit dem Zielkon-

---

ment gibt und […] für welchen Zweck das Experiment eingesetzt werden kann"; ähnl. *Springer*, SdW 2012, 9/48, 49: „Jedes Modell, so komplex es auch sein mag, ist weniger kompliziert als die Wirklichkeit, der es gerecht zu werden versucht – wahrscheinlich unendlich viel weniger." und weiter 52: „Doch daraus folgt nicht notwendig, dass jede Reduktion von Komplexität weniger real sein muss. Sie kann mehr oder weniger angemessen sein."

[136] *Arlen/Talley*, Experimental 2008, xx: "one must pay significant attention to whether an experiment captures the essential features of a real-world situation of interest." und weiter xx-xii: "there is no 'one size fits all' set of desiderata"; *Dipboye/Flanagan*, Am Psy'st 1979, 141, 149: "Rather than asking, Is this study externally valid?, a more appropriate question is, To what actors, settings, and behaviors may we generalize the findings of this study?"

[137] *Krauss*, Kugel 1998, 14.

[138] Allg. *DellaVigna*, J Econ Lit 2009, 315, 365: "Most phenomena that are important in laboratory experiments also affect decisions in a variety of economic settings."; skeptisch *Mitchell*, Persp Psy Sci 2012, 109, 115: "Applied lessons are often drawn from laboratory research before any crossvalidation work has occurred, yet […] laboratory findings may turn out to be [.] misleading about the nature of relations among variables outside the laboratory."

[139] *Fraidin*, UC Davis L Rev 2004, 1, 77: "Such claims must include more than simple identification of differences between settings. They must explain why these differences render the conceptual links invalid."

text abnimmt.[140] Versuchsteilnehmer suchen zur Handlungsorientierung ohnehin nach ähnlichen Situationen in ihrer früheren Erfahrung, daher wird der Versuchsleiter umso weniger über die Situationswahrnehmung der Teilnehmer wissen, je stärker er den Laborkontext durch Abstraktion verfremdet.[141] Das lässt sich allerdings nicht *nur* durch augenscheinliche Ähnlichkeit vermeiden. Oftmals wichtiger als die stumpfe Nachahmung der Wirklichkeit (*mundane realism*) – wie sie sogar von Rezipienten befürwortet wird, die ausdrücklich auch Grundlagenforschung zur empirischen Rechtsforschung zählen[142] – ist die Erzeugung einer eigenen experimentellen Wirklichkeit (*experimental realism*).[143] Damit ist gemeint, dass die Teilnehmer eines Experiments den Studienkontext genauso ernst nehmen müssen wie Entscheidungen in ihrem sonstigen täglichen Leben. Betrachtet man etwa die Handlungskonsequenzen, die den Teilnehmern der Studie von Ariely und Kollegen vor Augen standen, so lässt sich mit einiger Berechtigung fragen, ob für sie die Versuchssituation nicht viel realer wirkte als ihr „reales" Leben außerhalb des Versuchslabors.

Handlungskonsequenzen – also Anreize – sind nicht der einzige,[144] aber ein wichtiger und vieldiskutierter Aspekt experimenteller Wirklichkeit, deshalb gehe ich darauf nun noch näher ein.

b) Insbesondere: Anreizstrukturen

Verhalten wird maßgeblich durch seine Konsequenzen bestimmt, daher müssen Studienteilnehmer die Anreizstruktur des Experiments ernst nehmen. Wie bereits erwähnt ist das Experiment von Ariely und Kollegen insoweit untypisch, denn die Anreize in Höhe mehrerer Monatsgehälter, die die Autoren ihren Teilnehmern bieten konnten, lassen kaum Zweifel daran,

---

[140] Beachte aber *Shadish/Cook/Campbell*, Causal Inference 2002, 360: "matching operations by surface similarity […] is superficial [… and] hampered by the fact that not everyone will agree about which components are prototypical. By itself, then, surface similarity rarely makes a persuasive case for external validity, despite its wide use."

[141] *Harrison/List*, J Econ Lit 2004, 1009, 1050: "In an attempt to ensure generality and control by gutting all instructions and procedures of field referents, the traditional lab experimenter has arguably lost control to the extent that subjects seek to provide their own field referents."; ähnl. *Diekmann*, Sozialforschung 2012, 628.

[142] *Strempel* in: Broda u.a., FS Wassermann 1985, 223, 233 = *Strempel* in: Schäffer, Rechtskultur 1987, 87, 93: „je *besser* Ergebnisse empirischer Rechtsforschung sind, dh je genauer sie die soziale Wirklichkeit wiedergeben, […]".

[143] *Abelson*, Principled Argument 1995, 135 m.w.N.; *Swieringa/Weick*, J Acc Res 1982, S56, 57, 79 ff.; *McDermott*, Ann Rev Polit Sci 2002, 31, 56 m.w.N.

[144] Andere bei *Dipboye/Flanagan*, Am Psy'st 1979, 141, 147: "Awareness of being a research subject, unidimensionality of dependent measures, shortened time perspective, and low expectations of accountability for action are among the ways that laboratory and field settings may differ from a target setting."

dass jene die experimentelle Aufgabe *als Wirklichkeit* wahr- und ernstgenommen haben. Regelmäßig sind Laborexperimente dagegen nicht in der Lage, ihren Teilnehmern so erhebliche finanzielle Anreize zu bieten. Daher bedarf es einiger genauerer Überlegungen dazu, welche Anreize überhaupt erforderlich sind. Dabei sind zwei verschiedene Arten von Anreizen denkbar: drohende Verluste und winkende Gewinne.

Verluste sind im Experiment immer nur in engen Grenzen umsetzbar. Anders als die Teilnahme am Leben ist die Teilnahme an Experimenten freiwillig. Teilnehmer lassen sich dafür oft nur gewinnen, wenn sie sich darauf verlassen können, dass sie – wie immer sie entscheiden – jedenfalls kein Geld verlieren können.[145] (Andernfalls rückten Entscheidungslabore auch gefährlich in die Nähe von Glücksspielstätten.) In Experimenten werden Verlustszenarien deshalb meist anders inszeniert als dadurch, dass Teilnehmer Geld aus eigener Tasche zahlen müssen:

Zum einen ist es üblich, Teilnehmern einen festen Geldbetrag bereits für ihr Erscheinen zu zahlen (*show-up fee*). Dieser Betrag gilt oft als Mindestgewinn für die Teilnehmer, der in jedem Fall behalten werden darf. Von diesem Brauch kann jedoch abgewichen werden, wenn das Experiment es erfordert – etwa wenn der Umgang mit Verlusten zu untersuchen ist. Geht man also davon aus, dass Teilnehmer mit Betreten des Labors diesen Mindestbetrag bereits zu ihrem Vermögen zählen, sind die Verluste nicht perfekt gedeckelt.

Eine stärkere Variante dieser Manipulation besteht darin, Teilnehmer zu Beginn des Experiments erfolgsabhängig dafür zu bezahlen, dass sie eine anspruchslose Routinearbeit (*real effort task*) durchführen. Eine beliebte solche Aufgabe ist es, bestimmte Buchstaben in einem Buchstabensalat oder Ziffern in einer Matrix zu zählen. Dadurch, dass solch reizlose Aufgaben meist allein wegen der finanziellen Belohnung erfüllt werden, sehen sich Teilnehmer auch berechtigt, die Bezahlung zu beanspruchen. Zudem erfolgt die Bezahlung oft erfolgsabhängig, was die mentale Verbuchung als „eigene Leistung“ verstärkt. Viel spricht deshalb dafür, dass die Bezahlung für solche Aufgaben unmittelbar als eigenes Vermögen wahrgenommen wird. Greift der Experimentator in der Folge auf dieses Vermögen zu, werden Verluste durchaus real empfunden.

Als dritter Weg, ohne solche Manipulationen anzuwenden, lassen sich Verluste verbal simulieren, indem Teilnehmern die hypothetische Konsequenz ihrer Handlungen mitgeteilt wird – in der Hoffnung, dass sie diese Mitteilung ähnlich ernst nehmen wie eine real erlebte Konsequenz. Da die-

---

[145] Vgl. auch *Camerer/Hogarth*, J Risk Uncert 1999, 7, 36: “it is generally difficult to impose losses or punishments on subjects for bureaucratic reasons – university committees that approve protocols involving human subjects strongly object to it”.

selbe Methode für Gewinne denkbar ist, komme ich in diesem Zusammenhang darauf zurück.

Gewinne können den Teilnehmern eines Experiments im Prinzip in beliebiger Höhe in Aussicht gestellt werden. Weil die Teilnahme an Experimenten freiwillig ist und empirische Forscher nicht nur solche Teilnehmer ansprechen wollen, die gern Experimentalaufgaben lösen, ist die Zahlung einer Aufwandsentschädigung[146] ebenso praktisch notwendig wie die Begrenzung der Verluste. Streit entzündet sich jedoch an der Frage, ob Teilnehmern zusätzliche verhaltensabhängige Anreize gegeben werden müssen (sog. Inzentivierung), um sicherzustellen, dass sie jede ihrer Entscheidungen im Labor genauso ernst nehmen wie die Entscheidungen im Leben außerhalb. Soweit Anreize erforderlich sind oder jedenfalls geboten werden, stellt sich zudem die Frage nach deren Höhe.

Die erste Frage – Gibt es eine experimentelle Wirklichkeit ohne verhaltensabhängige Gewinnanreize? – ist eine der am heftigsten umkämpften Streitfragen zwischen vielen Ökonomen („Nein.") und den meisten Psychologen („Ja."[147]). Während die ablehnende Position mit dem ökonomischen Rationalmodell des Menschen als nutzenmaximierendem, arbeitsscheuem Eigenbrötler (*homo oeconomicus*) begründet wird, nehmen Befürworter an,

> „dass Menschen zu berichten fähig sind, was sie denken und in den beschriebenen Umständen tun würden, dass sie diese Informationen wahrheitsgemäß berichten wollen und dass dieser Wunsch hinreichend stark ist, um sie zur Aufgabenlösung nach besten Kräften anzuhalten."[148]

Diese Annahmen lagen vielen Jahrzehnten psychologischer Forschung zugrunde und beruhen durchaus nicht auf abwegigen Überlegungen:

Erstens ist bekannt, dass Menschen im Leben wie im Labor inneren (intrinsischen) Motivationen folgen, die durch äußere (extrinsische) Anreize ausgetrieben werden können (*crowding out*).[149] Gerade für aktienrechtliche Überwachungsorgane wird bisweilen angenommen, dass sie vorrangig nicht-monetären Anreizen folgen,[150] daher drohen inzentivierte Experimente, einen wichtigen Faktor menschlichen Entscheidungsverhaltens (die

---

[146] Das muss kein Geldbetrag wie die o.g. *show-up fee* sein: Die Teilnehmer an psychologischen Experimenten erhalten oft universitäre Leistungsgutschriften (Versuchspersonenstunden, VP-Stunden) oder kleine Sachgeschenke.

[147] Ausnahme etwa *Hertwig/Ortmann*, Behav Brain Sci 2001, 383 (mit 34 Repliken ab S. 403 und Duplik ab S. 433).

[148] *Read*, J Econ Meth 2005, 265.

[149] *Deci/Koestner/Ryan*, Psy Bull 1999, 627 (Metastudie über 128 Primärstudien); vgl. noch unten bei Fn. 250 und 270.

[150] *Cox/Munsinger*, L Ctmp Prob 1985, 3/83, 93: "The reward which dominates the thinking of […] a corporation's board of directors is nonmonetary."; vgl. auch unten Fn. 253.

„weiche" nicht-monetäre Motivation) auszublenden. *Zweitens* geht es in der empirischen Forschung nicht immer nur um rational reflektierte, sondern gerade auch um spontane Entscheidungen, die der rationalen Reflektion vorgelagert und einer Inzentivierung im engeren Sinne damit überhaupt nicht zugänglich sind.[151] *Drittens* ist der Übertragbarkeit von Erkenntnissen mitunter mehr geholfen, wenn Teilnehmer in hypothetischen Situationen entscheiden, die dem Zielkontext stark ähneln als in realen, die dem Zielkontext nur noch entfernt ähneln.[152] *Viertens* ist zumindest für Entscheidungen unter Risiko belegt, dass schon die Art, das Risiko sprachlich darzustellen (*framing*), psychologisch sehr ernst genommen wird.[153] Warum das für Gewinne und Verluste anders sein sollte, ist nicht einsichtig.

Letztlich nutzt auch die experimentelle Ökonomik Methoden, die der verbalen Simulation von Anreizen nahestehen, insbesondere die sog. Strategiemethode (*strategy method*),[154] die Versuchsteilnehmer nicht direkt miteinander interagieren lässt, sondern von jedem Teilnehmer abfragt, wie er auf jede mögliche Aktion seines Gegenübers reagieren will. Dieses Vorgehen wirft vollständigere Daten ab, sieht sich aber dem Einwand ausgesetzt, dass reales Entscheidungsverhalten „in der Hitze des Gefechts" stattfindet, und nicht strategisch „mit kühlem Kopf" (*hot vs. cold decision making*). Ein erster Literaturbericht kommt allerdings zu dem Ergebnis, dass unter 29 bisher publizierten Methodenvergleichen qualitativ durchweg die gleichen Ergebnisse erzielt werden und dass Unterschiede nur quantitativ und nur hinsichtlich bestimmter Verhaltensformen auftreten.[155] Auch der ökonomischen Experimentalpraxis sind hypothetische Entscheidungen also nicht ganz fremd. Der Streit muss (und kann) hier nicht näher aufgerollt oder gar entschieden werden, denn für die Frage der Übertragbarkeit einer Erkenntnis kommt es ohnehin auf den konkreten Versuchsplan der zugrundeliegenden Studie an.

Die zweite Frage im Zusammenhang mit den Gewinnanreizen – Wenn, dann in welcher Höhe? – ist weniger umstritten, aber nicht minder schwie-

---

[151] *Camerer/Hogarth*, J Risk Uncert 1999, 7; allenfalls kann die Inzentivierung als Auslösereiz (*cue*) dienen.

[152] So wohl *Read*, J Econ Meth 2005, 265, 274: "it is likely that in many cases the best possible study will be a hypothetical one that relies on participants being able to report what they would do if placed in a situation that cannot be precisely reproduced in the laboratory."; vgl. auch unten bei Fn. 284 zur Vignettenstudie.

[153] *Kühberger*, Org Behav Hum Dec Proc 1998, 23 (Metastudie über 136 Primärstudien).

[154] Eingeführt durch *Selten* in: Sauermann, Wirtschaftsforschung 1967, 136.

[155] *Brandts/Charness*, Exp Econ 2011, 375, 395: "while we do not claim that our study is definitive, it should at least dispel the impression that the strategy method inevitably yields results that differ significantly from results gathered using the traditional direct-response method. More research is needed to provide even more definitive evidence on this very pertinent issue in experimental economics."

rig. Einerseits kann die *zu geringe* Bemessung von Anreizen dazu führen, dass Teilnehmer weniger Motivation aufbringen als ganz ohne Bezahlung,[156] denn was als Gefälligkeit gern kostenlos getan wird, hätte als Dienstleistung oft einen hohen Preis.[157] Andererseits kann auch die *zu hohe* Bemessung der Anreize wirkungslos bleiben oder sogar unerwünschte Negativeffekte haben.[158] Dies sind letztlich alles empirische Fragen, und bisher existieren kaum aussagekräftige Forschungssynthesen. Vier Metastudien liefern zwar erste Anhaltspunkte, davon sind allerdings drei über zehn Jahre alt und methodisch teilweise fragwürdig:

1998 erschien eine Metastudie über 39 Primärstudien, von denen 32 unter Laborbedingungen stattgefunden hatten.[159] In diesen Studien hatte die *Existenz* finanzieller Anreize durchaus einen Einfluss auf die Leistung, wenn auch nur in quantitativen Aufgaben (also für die Leistungs*intensität* statt -qualität) und am schwächsten unter den Laborexperimenten.[160] Über die Auswirkungen der *Höhe* finanzieller Anreize konnte die Metastudie dagegen nichts sagen.[161]

1999 erschien ein Forschungsbericht, der freimütig eine willkürliche Auswahl von (ökonomischen) Primärstudien betrachtete.[162] Unter diesen 74 Studien konnten die meisten nicht belegen, dass die *Höhe* der Anreize einen nennenswerten Einfluss auf die durchschnittliche Leistung der Teilnehmer gehabt hätte[163] – und wenn, dann nur „in einfachen Aufgaben, die

[156] *Ariely/Gneezy u.a.*, Rev Econ Stud 2009, 451: “in some situations, paying a small amount in comparison to paying nothing seems to change the perceived nature of the task, which, if the amount of pay is not substantial, may result in a decline of motivation and effort” m.w.N.

[157] Anschaulich *Ariely*, Predictably Irrational 2008, 67 (Kap. 4, “Why We Are Happy to Do Things, but Not When We Are Paid to Do Them”).

[158] So ja das Ergebnis der Beispielstudie (bei Fn. 113) und der Forschung über *crowding-out* (bei und in Fn. 149).

[159] *Jenkins/Mitra u.a.*, J Appl Psy 1998, 777, 780 (eigene Auszählung der Tabelle).

[160] *Jenkins/Mitra u.a.*, J Appl Psy 1998, 777, 783: “the effect size is estimated to be .34 […] financial incentives do not affect performance quality, although this finding should be viewed with caution because it is based on only six studies. […] The strongest relationships were observed in experimental simulations (.56), followed by field experiments (.48) and laboratory experiments (.24).”

[161] *Jenkins/Mitra u.a.*, J Appl Psy 1998, 777, 784: “Although most studies contained information about incentive size, this information could not be coded easily in a way that made between-study comparisons meaningful.”

[162] *Camerer/Hogarth*, J Risk Uncert 1999, 7, 11: “Because of the opportunistic sampling we used, the reader is entitled to regard the paper as an informed essay or collection of conjectures, which may or may not prove true after a more careful meta-analysis of studies (and further research).”

[163] *Camerer/Hogarth*, J Risk Uncert 1999, 7, 30: “The most common result is no effect on mean performance.”

auf Anstrengung ansprechen […] Anreize schaden mitunter, wenn Probleme zu schwierig sind oder wenn einfache Intuitionen oder Gewohnheiten die beste Antwort vorgeben, während angestrengteres Nachdenken die Sache verschlimmert.“[164] Zumindest hinsichtlich der *Existenz* von Anreizen meinte der Literaturbericht zwar einen Einfluss festzustellen, konnte das aber nicht genauer quantifizieren.[165]

2000 schließlich folgte eine weitere, sorgfältiger durchgeführte Forschungsauswertung über 85 Primärstudien mit 131 Laborexperimenten.[166] Sie kam zu dem Ergebnis, dass die *Existenz* von Anreizen in 30–55 % aller Fälle die Leistung steigert, in 3–10 % aller Fälle die Leistung vermindert und in 42–60 % aller Fälle keine Auswirkungen hat.[167] In der weiteren Differenzierung zwischen verschiedenen Aufgabenarten zeigte sich erneut, dass die Anreizwirkung mit steigender Schwierigkeit der Aufgabe abnahm.[168] Daraus folgerte die Studie, dass „der Nutzen aus der Bezahlung von Teilnehmern untrennbar verbunden ist mit der Aufgabenart, die untersucht wird (einschließlich der Fähigkeiten des Teilnehmers) sowie vom gewählten Auszahlungssystem.“[169] Über die Anreiz*höhe* dagegen konnte die Forschungsauswertung nichts aussagen.[170]

2010 erschien die jüngste Metastudie, über 46 Laborexperimente. Sie untersuchte überhaupt nur die *Existenz* von Anreizen, nicht aber deren *Höhe*, und kam zu dem Ergebnis, dass leistungsabhängige Bezahlung die Leistung nur in solchen Aufgaben steigere, die als langweilig oder uninteressant gelten, während sie in interessanten Aufgaben die Leistung sogar verringere.[171]

Den vier zitierten Forschungssynthesen gemein ist also die Erkenntnis, dass je interessanter oder kognitiv anspruchsvoller die dem Teilnehmer gestellte Aufgabe ist, desto weniger finanzielle Anreize dazu beitragen, eine

---

[164] *Camerer/Hogarth*, J Risk Uncert 1999, 7, 34.

[165] *Camerer/Hogarth*, J Risk Uncert 1999, 7, 21: “a disproportionate number of effects result from raising the level of incentives from 0 […] This suggests that while adding some incentive to otherwise-hypothetical choices often matters, experiments which then multiply stakes by 2, 4, or 20 do not produce similar boosts in performance.”

[166] *Bonner/Hastie u.a.*, J Mgmt Acc Res 2000, 19 mittels quantitativer, aber nicht metaanalytischer Methoden (ebd. 32).

[167] *Bonner/Hastie u.a.*, J Mgmt Acc Res 2000, 19, 34 f.: Schätzungen jeweils für *best* und *worst case* (erläutert ebd. 29 f.).

[168] *Bonner/Hastie u.a.*, J Mgmt Acc Res 2000, 19, 38: “as tasks become more cognitively complex […] it is less likely that financial incentives will improve performance.”

[169] *Bonner/Hastie u.a.*, J Mgmt Acc Res 2000, 19, 39.

[170] *Bonner/Hastie u.a.*, J Mgmt Acc Res 2000, 19, 40: “for a number of reasons we were not able to incorporate this variable into our analysis.”

[171] *Weibel/Rost/Osterloh*, J Publ Admin Res Theory 2010, 387, 397: “Pay for performance increases performance in the case of noninteresting tasks (0.42***), whereas in the case of interesting tasks pay for performance reduces performance (-0.13***).”

vom Teilnehmer ernstgenommene experimentelle Wirklichkeit zu erzeugen. Wie sich die Höhe dieser Anreize auswirkt, lässt sich der bisherigen Forschung nicht allgemein entnehmen. Allerdings sind in den mehr als zehn Jahren seit Erscheinen der drei Forschungssynthesen, die auch die Anreizhöhe untersuchten, zahlreiche neue Primärstudien erschienen, die bisweilen (wie Ariely und Kollegen) ganz erhebliche Anreizhöhen einsetzten. Ob und in welcher Höhe finanzielle Anreize in Laborexperimenten erforderlich und hilfreich sind, ist daher eine von zwei Fragen, für die kürzlich das akute Bedürfnis nach einer Metastudie artikuliert wurde.[172]

c) Homogenität der Versuchsteilnehmer

Eine dritte methodische Beschränkung von Experimenten ist die soziale Homogenität (Übereinstimmung) ihrer Versuchsteilnehmer. Diese Homogenität fördert zwar die innere Gültigkeit, denn Unterschiede zwischen den Experimentalgruppen lassen sich leichter auf die experimentelle Manipulation zurückführen statt auf vielfältige Unterschiede zwischen den teilnehmenden Personen.[173] Andererseits geht gerade diese Homogenität auf Kosten der Übertragbarkeit und vermindert dadurch die äußere Gültigkeit. Wie die meisten Laborstudien, so erfolgten auch zwei der drei Experimente von Ariely und Kollegen mit studentischen Versuchsteilnehmern. Anders als die meisten Laborstudien führten Ariely und Kollegen ihren ersten Versuch allerdings mit indischen Dorfbewohnern durch, und belegten dadurch in gewissem Umfang die Übertragbarkeit ihrer Erkenntnisse. In der üblichen Experimentalpraxis hingegen sind die Teilnehmer in zweierlei Hinsicht homogen, die Übertragbarkeit also beschränkt:

Zum einen stammen die Versuchsteilnehmer meist aus dem gleichen Kulturraum, nämlich aus demjenigen, den ein vielbeachteter Forschungsbericht mit dem Wortspiel "WEIRD" charakterisierte.[174] Das englische Adjektiv *weird* bedeutet „eigenartig", lässt sich aber zugleich als Akronym verstehen für fünf Adjektive, die diesen eigenartigen Kulturraum beschreiben: westlich (*Western*), gebildet (*Educated*), industrialisiert (*Industrialized*), reich (*Rich*) und demokratisiert (*Democratic*). Der Forschungsbericht

---

[172] *Feltovich*, J Econ Surv 2011, 371, 372, 375 f. (zur anderen unten bei Fn. 187); vgl. auch schon Fn. 162.

[173] *Peterson*, J Cons Res 2001, 450, 453: "the alleged homogeneity of college students translates into stronger hypothesis tests than if nonstudents were the research subjects since there is less noise or extraneous variation associated with college students than with nonstudents. Indeed, this assumption is frequently provided as the rationale enabling the use of college students as research subjects."; meta-analytisch bestätigt auf S. 458.

[174] *Henrich/Heine/Norenzayan*, Behav Brain Sci 2010, 61 (mit 28 Repliken ab S. 83 und ausf. Duplik ab S. 111); Zusammenfassungen etwa in *Henrich/Heine/Norenzayan*, Nature 2010, 29; *Jones*, Sci 2010, 1627.

kritisiert, dass beispielsweise 96 % der Studien, die 2003–2007 in den renommiertesten psychologischen Zeitschriften veröffentlicht wurden, mit Teilnehmern aus westlichen Industrienationen (also 12 % der Weltbevölkerung) stattfanden, obwohl dieselben Studien vorgeben, *menschliches* Verhalten im Allgemeinen zu erforschen.[175] Dadurch werde völlig überspielt, welches menschliche Verhalten tatsächlich universell ist (weil es etwa auf evolutionären Grundfunktionen beruht) und welches durch Ethnie, Kultur oder Sozialisierung spezifisch geprägt ist. Dabei treten mitunter überraschende Erkenntnisse zutage. Sogar so vermeintlich fundamentale menschliche Fähigkeiten wie räumliche oder visuelle Wahrnehmung variieren zwischen den Kulturen:[176] Der berühmten Müller-Lyer-Täuschung etwa,[177] der sich in westlichen Gesellschaften kaum jemand entziehen kann, unterliegen südafrikanische Minenarbeiter oder die San-Buschleute der Kalahari praktisch gar nicht – sie nehmen die Müller-Lyer-Linien tatsächlich als gleich lang wahr. Die Übertragbarkeit von empirischen Studien auf Menschen anderer Ethnien, Kulturen oder Sozialisierung muss also im Einzelfall sorgfältig geprüft werden,[178] zumal selbst innerhalb der westlichen Gesellschaften deutliche kulturelle Unterschiede festzumachen sind.[179] Auch zur Vergütungspraxis liegt bereits eine interkulturelle Vergleichsstudie vor.[180]

Zum anderen stammen die Versuchsteilnehmer ganz überwiegend aus den Anfangssemestern der Universitäten, repräsentieren also eine Gruppe, die sogar innerhalb ihrer eigenen Gesellschaft sehr eigenartig und homogen erscheint: tendenziell jünger, besser gebildet und mit einem typischen Persönlichkeitsbild, das sich von demjenigen anderer Personenkreise deut-

---

[175] *Henrich/Heine/Norenzayan*, Behav Brain Sci 2010, 61, 63, und weiter 79: "WEIRD people are […] one of the worst subpopulations one could study for generalizing about *Homo sapiens*." (ebenso 82).

[176] *Henrich/Heine/Norenzayan*, Behav Brain Sci 2010, 61, 64 f. (visuelle Wahrnehmung), 68 (räumliche Wahrnehmung), je m.w.N., vgl. auch 78: "Many of the processes […] that vary dramatically across populations would seem to be 'basic' psychological processes."; anders wohl noch *Swieringa/Weick*, J Acc Res 1982, S56, 69: "Experiments may be better suited for studying cognition than for studying either sentiment or action."

[177] Entdeckt 1889 von *Franz Müller-Lyer*: Eine von spitzen Winkeln eingeschlossene gerade Linie erscheint je nach Richtung der Winkelspitze unterschiedlich lang (Abb. bei *Henrich/Heine/Norenzayan*, Behav Brain Sci 2010, 61, 64).

[178] *Henrich/Heine/Norenzayan*, Behav Brain Sci 2010, 61, 79: "At this point, we know of no strong grounds to make a priori claims to the 'fundamentalness' or the likely universality of a given psychological process."

[179] *Henrich/Heine/Norenzayan*, Behav Brain Sci 2010, 61, 76: "American participants are exceptional even within the unusual population of Westerners – outliers among outliers." (m.w.N. ab 74).

[180] *Merchant/Van der Stede u.a.*, Eur Acc Rev 2011, 639 m.w.N.

lich unterscheidet.[181] Wiewohl diese Feststellung einer der häufigsten Kritikpunkte an der empirischen Laborforschung ist und zu verschiedenen Zeiten in ganz unterschiedlichen Disziplinen geäußert wurde,[182] waren entsprechende empirische Erkenntnisse lange sehr verstreut und kaum systematisch ausgewertet. Erst 2001 verglich ein Forschungsbericht die Ergebnisse von 34 Metastudien unter dem Gesichtspunkt, ob Studenten sich systematisch anders verhielten als nichtstudentische Teilnehmer.[183] Der Forschungsbericht konnte dabei zwar nicht zwischen Laborexperimenten und anderen Untersuchungen unterscheiden,[184] kommt aber zu dem Ergebnis, dass in 81 % der untersuchten Fragestellungen *qualitativ* gleiche Ergebnisse für Studenten und Nichtstudenten herauskommen, dass aber nur 52 % der untersuchten Fragestellungen auch *quantitativ* ähnliche Ergebnisse erbringen.[185] Was die verbleibenden Unterschiede angeht, konnte der Forschungsbericht keine systematischen Ursachen oder Muster ausmachen.[186] Daher überrascht es nicht, dass die überspitzte Frage „ob Universitätsstudenten auch Menschen sind“ eine von zweien ist, für die kürzlich das akute Bedürfnis nach einer Metastudie artikuliert wurde.[187] Einstweilen lässt sich nur feststellen, dass die mit Studenten gewonnen Ergebnisse zwar qualitativ in gleicher Weise bei anderen Personengruppen auftreten, dass aber die jeweilige Effektgröße stark variieren kann.[188]

Der Befund ist ermutigender als er zunächst scheint. Laborexperimente dienen dank ihrer regelmäßig hohen inneren Gültigkeit dazu, Ursachenbe-

---

[181] *Sears*, J Pers Soc Psy 1986, 515, 521, 527: “students tend, among other things, to have incompletely formulated senses of self, rather uncrystallized sociopolitical attitudes, unusually strong cognitive skills, strong needs for peer approval, tendencies to be compliant to authority, quite unstable group relationships, little material self-interest in public affairs, and unusual egocentricity.”; vgl. auch *Henrich/Heine/Norenzayan*, Behav Brain Sci 2010, 61, 76 f.

[182] *Peterson*, J Cons Res 2001, 450 m.w.N.: “there seem to be few insights into the college-student-as-research-subject issue that have not been previously presented, replied to, subjected to rejoinder, and subsequently forgotten.”

[183] *Peterson*, J Cons Res 2001, 450 (4 Metastudien über ca. 650.000 Teilnehmer und weitere 30 Metastudien über ca. 350.000 Teilnehmer).

[184] *Peterson*, J Cons Res 2001, 450, 8 Fn. 2.

[185] *Peterson*, J Cons Res 2001, 450, 455: “the directionality of the relationships investigated differed between college students and nonstudents 19% of the time. […] in an additional 29% of the relationships investigated, the larger effect size in a pair exceeded the smaller one by a factor of more than two.”; krit. *Trottier/Gordon*, SSRN 2011, 5: “However, this does not necessarily mean that different conclusions would be reached.”

[186] *Peterson*, J Cons Res 2001, 450, 458: “there was no systematic or unambiguous pattern to the differences.”

[187] *Feltovich*, J Econ Surv 2011, 371, 372 (Zitat dort), 374; andere Frage oben bei Fn. 172.

[188] Vgl. schon *Sears*, J Pers Soc Psy 1986, 515, 520: “the nature of the relationship may be wrongly described [… but] more likely is that the strength of the relationship may be wrongly described.”

ziehungen festzustellen und elementare Mechanismen aufzuklären. Ob oder ob nicht eine bestimmte Ursache (z.B. die Vergütungshöhe) eine Rolle (z.B. in der Arbeitsmotivation) spielt, ist oft die entscheidende Frage, die das Experiment helfen soll, zu beantworten. Festzustellen, *wie groß* diese Rolle ist, ist dagegen selten Anspruch der experimentellen Arbeit, die oft „nur“ als Minimaltest fungiert.[189] Zudem ist die Feststellung, dass die Erkenntnisse über Studenten auf Nichtstudenten oft nicht übertragbar sind, noch recht grob. Die Gruppe der Nichtstudenten ist ja selbst sehr durchmischt und dem relevanten Zielkontext nicht unbedingt ähnlicher als die Gruppe der Studenten. Letztlich kommt es für Fragen der Vorstandsvergütung darauf an, wo Studenten im Vergleich zu *Vorständen* (nicht Nichtstudenten allgemein) in den fünf Dimensionen der Übertragungslehre (oben bei Fn. 22) stehen. Beispielsweise lässt sich einerseits argumentieren, dass sie die vergleichsweise weniger gefestigten Persönlichkeiten und weniger erfahrenen Entscheidungsträger sind, andererseits sind sie hinsichtlich ihres Bildungshintergrunds, der intellektuellen Fähigkeiten und Leistungsmotivation oft mit Unternehmensführern vergleichbar.[190]

Zwei der eben angesprochenen Vergleichskriterien – Bildung und Erfahrung – sind Teilaspekte der allgemeinen professionellen Expertise, auf die ich nun abschließend eingehe.

d) Insbesondere: Expertisemangel

Die bevorzugte Rekrutierung von Studenten für die Teilnahme an Laborexperimenten wird vor allem und immer dann besonders kritisch gesehen, wenn die so gewonnen Ergebnisse auf Personenkreise übertragen werden sollen, die über eine höhere Expertise (also bessere Ausbildung oder größere Erfahrung) mit den zu treffenden Entscheidungen verfügen. Vergleichsstudien zwischen Experten- und Laienurteilen wurden etwa mit Gerichtssachverständigen,[191] Wirtschaftsprüfern[192] und Unternehmensleitern[193]

---

[189] *Croson*, U Ill L Rev 2002, 921, 942: “if it does not work in the lab, it is unlikely to work in reality.”; pointiert *Ashton/Kramer*, J Acc Res 1980, 1, 13: “students to the behavioral researcher may be like fruit flies to the geneticist”.

[190] *Croson*, U Ill L Rev 2002, 921, 943: “There is no reason to think, in general, that students make decisions differently than professionals […] This is highlighted by the fact that the students of today are the professionals of tomorrow.”; *Trottier/Gordon*, SSRN 2011, 5: “the justification that business students are future managers is particularly compelling.” m. Verw. auf *Remus*, J Bus Res 1986, 19; ebenso *Remus*, Manage Dec Econ 1996, 93, 94.

[191] *Dror/Rosenthal*, J Forensic Sci 2008, 900.

[192] *Bédard*, Acc Org Soc 1989, 113; *Smith/Kida*, Psy Bull 1991, 472, 487: “heuristics and biases often observed in studies using subjects with little or no expertise […] were also evident to some extent in the judgments of experienced professional auditors.”

[193] *Remus*, Manage Dec Econ 1996, 93, 94 f. m.w.N.; *Trottier/Gordon*, SSRN 2011, 6 m.w.N.; *Bolton/Ockenfels/Thonemann*, Mgmt Sci 2012, 2225 m.w.N.

durchgeführt. Sie kommen meist zu dem Ergebnis, dass die jeweilige Expertise einen deutlich geringeren Einfluss auf wichtige Aspekte menschlichen Verhaltens hat (etwa auf die Risikowahrnehmung) als gemeinhin angenommen.[194] Expertise kann sich sogar negativ auswirken, wenn sie dazu führt, dass sonstige Entscheidungshilfen ignoriert und deshalb weniger kompetente Entscheidungen getroffen werden.[195]

Die für die vorliegende Arbeit interessanteste Expertengruppe sind Unternehmensleiter. Ob sie sich dank ihrer Expertise anders verhalten als Laien (bzw. Studenten),[196] ist bisher nur bruchstückhaft erforscht. Aufgrund der ganz praktischen Erwägung, dass Unternehmensleiter sich nicht so leicht für Experimente rekrutieren lassen wie Studenten,[197] sind entsprechende Studien bisher nicht allzu zahlreich. Forschungssynthesen fehlen deshalb. Für die Frage, inwieweit der studentische Expertisemangel im Vergleich zu Unternehmensleitern die Übertragbarkeit der Erkenntnisse beschränkt, können deshalb nur erste Hinweise aus einigen Primärstudien entnommen werden.

Eine Studie von 1986 untersuchte die (hypothetischen) Produktionssteuerungsentscheidungen der Teilnehmer eines (wohl in den USA durchgeführten) MBA-Abendkurses.[198] Zwischen dem Viertel der Teilnehmer mit der meisten Management-Erfahrung – alle mindestens aus der zweiten Hierarchieebene ihres Unternehmens – und dem Viertel der Teilnehmer mit abgeschlossenem Studium, aber ohne Management-Erfahrung fand die Studie keine systematischen Unterschiede im Entscheidungsverhalten.[199] We-

---

[194] *Rowe/Wright*, Risk Analysis 2001, 341 (Forschungsbericht über 9 Studien zur Risikowahrnehmung); prägnant *Bostrom*, Duke Env L Policy F 1997, 101, 113: "Comparisons of expert risk assessments with lay risk perceptions can all too easily focus on differences in knowledge and expertise, while ignoring equally real differences in individual contexts, motives, and even values."

[195] *Arkes/Dawes/Christensen*, Org Behav Hum Dec Proc 1986, 93, 107: "such 'experts' performed worse but had higher confidence in their performance than the nonexperts."; vgl. auch unten § 4 C.V.1. zur Selbstüberschätzung von Unternehmensleitern.

[196] *Remus*, J Bus Res 1986, 19, 20: "not unexpectedly, managers and students often have different attitudes. […] There is, however, general agreement that attitudes are not necessarily related to behavior (e.g., decisions actually made)."; vgl. schon *Abdel-Khalik*, Acc Rev 1974, 743, 744 und *Ashton/Kramer*, J Acc Res 1980, 1 ff., je m.w.N.

[197] *Homberg/Osterloh*, JfB 2010, 269, 275: „Managerinnen und Manager stellen sich üblicherweise nicht für psychologische Tests zur Verfügung."; *Abdel-Khalik*, Acc Rev 1974, 743: "reluctance of many executives to provide the needed cooperation"; *Remus*, Manage Dec Econ 1996, 93, 94: "student subjects are readily available in an academic setting while it is difficult to get managers to be subjects."

[198] *Remus*, J Bus Res 1986, 19, 21; MBA steht für *master of business administration*; wo der Kurs stattfand, ist nicht berichtet, aber der Autor ordnete sich der University of Hawaii zu.

[199] *Remus*, J Bus Res 1986, 19, 23: "It appears that in tasks such as production scheduling, MBA students with little business experience can be safely used as surrogates for managers."

nige Jahre später erweiterte derselbe Autor seine Studie allerdings um Studenten ohne abgeschlossenes Studium und stellte fest, dass deren Entscheidungen sich in der Tat deutlich von den MBA-Studenten (mit und ohne Management-Erfahrung) unterschieden.[200] Gleichwohl räumte der Autor ein, dass auch Studenten ohne Abschluss zumindest *qualitativ* dieselben Verhaltensmuster aufweisen könnten wie Unternehmensleiter.[201]

Eine deutlich jüngere Studie befragte kanadische Unternehmensleiter sowie Studenten im vierten Studienjahr mit einem abgeschlossenen Kurs in Buchhaltungstheorie.[202] Beide Gruppen sollten die Entscheidung eines fiktiven Unternehmensleiters vorhersagen, der entscheiden muss, inwieweit er ein derzeit überbewertetes Wirtschaftsgut abschreibt oder auf eine Werterholung hofft. Auch diese Studie kam zu dem Ergebnis, dass die Studenten zwar *quantitativ* systematisch andere Angaben machten als die Unternehmensleiter, dass ihr Verhalten aber zumindest *qualitativ* ähnlich genug war, um die daraus gewonnenen Erkenntnisse auf die Unternehmensleiter zu übertragen.[203] Dabei bildeten Studenten mit bestimmten Merkmalen eine bessere Annäherung als andere,[204] nämlich diejenigen mit Arbeitserfahrung und einem wirtschaftsorientierten Abendkurs sowie diejenigen mit mittelmäßigen statt sehr guten Noten in Buchhaltungstheorie.

Eine dritte Studie verglich die Entscheidungen von deutschen Studenten mit und ohne ersten Studienabschluss mit denjenigen von Einkäufern der mittleren Managementebene.[205] Beide sollten wiederholt über Vorratsbeschaffungen bei schwankender Nachfrage entscheiden und wurden nach der Profitabilität ihrer Vorratshaltung bezahlt. Während die erfahrenen Einkäufer anfangs bessere Entscheidungen trafen als die Studenten, holten diese nach Erhalt zusätzlicher Informationen auf, und zuletzt trafen die Absolventen sogar bessere Entscheidungen als die Berufseinkäufer.[206]

Die drei geschilderten Studien sind freilich nur illustrativ, deuten aber darauf hin, dass Studenten schon vor dem Studienabschluss qualitativ ähn-

---

[200] *Remus*, Manage Dec Econ 1996, 93, 98: “the undergraduates made significantly more costly decisions, had less effective heuristics, and were more erratic.”

[201] *Remus*, Manage Dec Econ 1996, 93, 99: “students […] might show the same treatment effects as managers.”

[202] *Trottier/Gordon*, SSRN 2011, 11.

[203] *Trottier/Gordon*, SSRN 2011, 3, 19: “while the students appraised a significantly lower likelihood that the impairment would be recorded, the pattern of their responses was similar enough to the manager responses to lead to the same inference and conclusions.”

[204] *Trottier/Gordon*, SSRN 2011, 20: “some students are stronger proxies for managers.”

[205] *Bolton/Ockenfels/Thonemann*, Mgmt Sci 2012, 2225.

[206] *Bolton/Ockenfels/Thonemann*, Mgmt Sci 2012, 2225, 2231 f.: “managers are no better than students in utilizing information relevant to rational inventory stocking […] In Phases 2 and 3 of all experiments, graduates tended to outperform managers and freshmen. […] students, particularly graduates, use information and training better than managers overall”.

liche Verhaltensmuster aufweisen wie Unternehmensleiter[207] und dass sich diese Ähnlichkeiten nach Abschluss des Erststudiums bzw. Aufnahme eines Aufbaustudiums auch quantitativ niederschlagen. Allerdings rekrutierten alle drei Studien ihre Teilnehmer aus wirtschaftswissenschaftlichen Studiengängen, daher ist es durchaus möglich, dass Studenten anderer Fachrichtungen sich schon qualitativ anders entscheiden; dafür sprechen etwa mehrere Studien, die deutliche Verhaltensunterschiede zwischen Studenten der Wirtschaftswissenschaften und denen anderer Fachrichtungen belegen.[208] Auch hier hängt letztlich viel davon ab, wie gut die jeweilige Studienplanung auf die Forschungsfrage abgestimmt ist.[209] Demnach lässt sich nicht pauschal sagen, ob die mit Studenten gewonnenen Erkenntnisse auf Unternehmensleiter übertragbar sind oder nicht, sondern es kommt vor allem auf die Einzelheiten der jeweiligen Studie an.[210]

Fragt man konkret, ob die Studienergebnisse von Ariely und Kollegen auf andere Personenkreise mit besserer Ausbildung und größerer Berufserfahrung übertragbar sind, so lassen sich Hinweise etwa aus einem Experiment mit 182 Wirtschaftsprüfern der KPMG entnehmen, die nach etwa drei Jahren Berufstätigkeit an einem Weiterbildungskurs teilnahmen.[211] Dort sollten die Teilnehmer das *Moody's*-Rating von 16 Unternehmensanleihen aufgrund von deren Finanzkennzahlen vorhersagen. Für einige der Teilnehmer wurde die Aufgabe dadurch erschwert, dass sie einen finanzmathematischen Algorithmus mit 50-prozentiger Zuverlässigkeit übertreffen mussten. In dieser kognitiv sehr anspruchsvollen Situation verminderte ein finanzieller Anreiz mit Wettbewerbskomponente abermals die Leistung.[212]

---

[207] *Bolton/Ockenfels/Thonemann*, Mgmt Sci 2012, 2225, 2233: "This seems to suggest that the use of student subjects is justified for analyzing the direction of the effects in manager populations but that manager subjects should be used for estimating the magnitude of effects in manager populations."

[208] Etwa *Rubinstein*, Econ J 2006, C1 ("the overly mathematical methods used to teach economics encourage students to lean towards profit maximisation."); *Ferraro/Pfeffer/Sutton*, Acad Mgmt Rev 2005, 8, 14; *Bauman/Rose*, J Econ Behav Org 2011, 318, 319, je m.w.N.

[209] *Trottier/Gordon*, SSRN 2011, 2: "with the careful planning and execution of experiments and with a judicious selection of students, we can produce results that can be generalized with confidence." m. Bsp. u.w.N.

[210] *Trottier/Gordon*, SSRN 2011, 2: "We are not likely to ever reach the sweeping conclusion that students are *always* or *never* suitable surrogates for managers. Rather, the literature should move in the direction of identifying attributes that lead to a strong correspondence between the student and manager responses" und weiter 20: "Researchers need to replicate a far greater number of experiments before we can fully understand the benefits and limitations of using students as proxies."; genauso schon *Abdel-Khalik*, Acc Rev 1974, 743, 750.

[211] *Ashton*, J Acc Res 1990, 148, 160.

[212] *Ashton*, J Acc Res 1990, 148, 166, 173: "When the [.] decision aid is available *and* the incentive is offered, […] there is decreased reliance on the aid, decreased mean accuracy, and increased variability across subjects".

Damit scheint der von Ariely und Kollegen belegte Effekt nicht allein auf der mangelnden Entscheidungsexpertise der Studenten zu beruhen. Allgemein sind mindestens vier Gründe dafür denkbar, dass viele laborexperimentelle Befunde auch in der Gegenwart größerer Expertise fortbestehen: Erstens erhalten Entscheider in der Realität selten eindeutige Rückmeldungen über die Auswirkungen ihres Handelns, zweitens kann zunehmende Erfahrung einzelne Verhaltenstendenzen noch verstärken, drittens fehlt Experten mitunter der Anreiz, ihre Expertise auch einzusetzen, und viertens beruhen viele Verhaltensmuster (z.B. Vorlieben) auf völlig expertiseunabhängigen Mechanismen.[213]

IV. Das Spektrum der Methoden

1. Von der Dichotomie zum Spektrum

Nachdem die beiden Idealtypen von Übertragbarkeit und Eindeutigkeit und ihre jeweiligen methodischen Beschränkungen vorgestellt sind, sollte dreierlei deutlich geworden sein:

1. Gültigkeit ist ein theoretisch beliebiges, aber praktisch hilfreiches Beurteilungskriterium für zwei wichtige Aspekte der Güte empirischer Forschung (dazu oben A.I.).
2. Gültigkeit ist eine Eigenschaft der empirischen Erkenntnis, nicht ihrer Gewinnungsmethode:[214] Experimente können (nicht müssen) hohe innere Gültigkeit auf Kosten der äußeren erzielen, Beobachtung kann (nicht muss) hohe äußere Gültigkeit auf Kosten der inneren erzielen.[215]
3. Abstraktion ist in der Gewinnung empirischer Erkenntnis prinzipiell unvermeidlich,[216] daher kann die Abwägung zwischen den Gültigkeiten nur anhand einer konkreten Untersuchungsfrage beurteilt werden.

Das konkrete Ergebnis der Abwägung zwischen den Gültigkeiten bestimmt letztlich die Methodenwahl.[217] Dadurch entstehen zwischen den Ex-

---

[213] *DellaVigna*, J Econ Lit 2009, 315, 365 f.

[214] *Shadish/Cook/Campbell*, Causal Inference 2002, 34: “Validity is a property of inferences. It is *not* a property of designs or methods”.

[215] *Dipboye/Flanagan*, Am Psy’st 1979, 141, 147: “there is no empirical basis for a belief in the inherent external validity of field research.”; *Swieringa/Weick*, J Acc Res 1982, S56, 86: “the field can be less real than the laboratory. […] The field merely reflects the final edited version of quite different processes […] that over time have become modified into what we now arbitrarily call the real world.”

[216] Siehe oben nach Fn. 122; in der kantischen Dichotomie (dazu, wenngleich ablehnend, *Upmeier*, Fakten 2010, 47) sind empirische Erkenntnisse *synthetisch* und deshalb abstraktionsbedürftig. Nur *analytische* Urteile können darauf verzichten, bringen dafür aber auch keine wirklich neue Erkenntnis.

trempolen von Beobachtung und Experiment zahlreiche Schattierungen.[218] So ist die Korrelationsstudie als Idealtyp der Übertragbarkeit nur eine von vielen Formen der Beobachtung. Ganz grob lassen sich *offene* und *verdeckte*, *teilnehmende* und *nicht-teilnehmende* sowie *Selbst-* und *Fremdbeobachtung* unterscheiden,[219] wobei die Korrelationsstudie als nicht-teilnehmende verdeckte Fremdbeobachtung zu qualifizieren wäre. Auch vom anderen Ende des Spektrums her haben sich aus dem Laborexperiment zahlreiche Varianten entwickelt – etwa Feldexperimente, Quasi-Experimente, natürliche Experimente, usw. – die letztlich fast nahtlos in die Kategorien der Beobachtung übergehen: Experimente lassen sich ebenso als besondere Form der Beobachtung verstehen,[220] wie sich Beobachtungen als besondere Formen des Experiments verstehen lassen.[221]

Daher ist es müßig zu fragen, ob die zwischen Laborexperiment und Korrelationsstudie denkbaren Abstufungen dazu dienen, die Übertragbarkeit des Laborexperiments oder die Eindeutigkeit der Korrelationsstudie zu erhöhen.[222] Historisch mag zwar ein Vorsprung der beobachtenden Methoden zu beobachten sein, weshalb viele Studien von der Beobachtungsform

---

[217] So schon *Häcker* in: Bender, Tatsachenforschung 1972, 143, 155: „Wird ein konkretes Untersuchungsvorhaben geplant, so wird vorher zu entscheiden sein, auf welche Vorteile man Wert legen will und welche Experimentiertechnik man demzufolge wählen muss.“; ebenso *Hussy/Schreier/Echterhoff*, Forschungsmethoden 2010, 134 (Illustration), 135: „Neben den gegebenen Rahmenbedingungen, die ein Wissenschaftler vorfindet, bedingt somit auch die Zielsetzung (Priorität für die interne oder externe Validität) die Auswahl der experimentellen Variante.“; *Lawless/Robbennolt/Ulen*, Methods 2010, 46: “for any given research issue, […] any empirical approach will have a different set of advantages and disadvantages and [...] degree to which they prioritize […] internal validity or generalizability”.

[218] *Henry*, Psy Inq 2008, 49, 66: “Too often, evidence for causality is interpreted such that either you have it (via experimentation) or you do not (nonexperimentation). However, evidence for causality instead should be thought of as existing on a continuum.”

[219] *Röhl*, Rechtssoziologie 1987, 109; *Beller*, Forschen lernen 2008, 33; sehr knapp *Lercher* in: Barta u.a., Rechtstatsachenforschung 2009, 205, 228 ff.

[220] *Diekmann*, Sozialforschung 2012, 548: „In einem allgemeinen Sinn sind sämtliche empirische Methoden Beobachtungsverfahren.“; *Lercher* in: Barta u.a., Rechtstatsachenforschung 2009, 205: „Experiment als ein Sonderfall der Laborbeobachtung“; ebenso *Häcker* in: Bender, Tatsachenforschung 1972, 143, 146; *Beller* (Fn. 31); *Sedlmeier/Renkewitz* (Fn. 229).

[221] *Hussy/Schreier/Echterhoff*, Forschungsmethoden 2010, 135: „Feldstudie als letzte ernst zu nehmende Sonderform des Experiments“; noch weiter *Harrison/List*, J Econ Lit 2004, 1009: “In some sense every empirical researcher is reporting the results of an experiment.”

[222] Vgl. *Holland*, J Am Stat Assoc 1986, 945, 954: “the notion of cause that operates in an experiment *and* in an observational study *is the same*. The difference is in the degree of control an *experimenter* has over the phenomena under investigation compared with that which an *observer* has.”; *Epstein/Martin* in: Cane/Kritzer, Hdb Empirical 2010, 901, 904: “other than issues of data generation and control […], experimental and observational studies are not altogether different”; mit *Diekmann*, Sozialforschung 2012, 329 ff. lassen sich die Methoden im Hinblick auf ihre statistische Auswertung als Varianten der *Varianzkontrolle* kennzeichnen.

ausgehen und sie dem Laborideal anzunähern versuchen.[223] Daher existieren vermutlich mehr und erprobtere Methoden, um innerhalb beobachtender Studien die Beschränkungen der inneren Gültigkeit zu verringern[224] als solche, die innerhalb experimenteller Methoden die Beschränkungen der äußeren Gültigkeit verringern.[225] Diese Entwicklungslinien unterscheiden sich aber zwischen den Disziplinen ganz erheblich und liegen keineswegs in der Natur der Sache begründet.

2. Die Schattierungen des Spektrums

Zur Illustration der methodischen Vielfalt und zur Erläuterung der feinen Abstufungen von innerer und äußerer Gültigkeit, mit denen verschiedene Methoden einhergehen, werden nun noch fünf Methoden aus dem Spektrum zwischen Korrelationsstudie und Laborexperiment vorgestellt.

Dabei bleibt eine wichtige empirische Studienform außen vor, die auch im Gesellschaftsrecht Verwendung findet,[226] wenngleich mir keine Studien zur Vorstandsvergütung bekannt sind: Die Dokumentenauswertung. Ihre quantitative Variante wird oft als „quantitative Inhaltsanalyse“[227] oder treffend als „Aktenuntersuchung“[228] bezeichnet, lässt sich aber zwanglos als Korrelationsstudie mit besonderer Datenquelle verstehen, daher betrachte ich sie nicht als eigenständige Methode der quantitativen Forschung.

Stattdessen orientiere ich mich für die Auswahl der nun darzustellenden Studienarten an zwei gebräuchlichen Kategorisierungen,[229] ergänzt um die

---

[223] *List*, Sci 2008, 207: “The fundamental challenge in the social sciences is how to go beyond correlational analysis to provide insights on causation.”; *Ho/Rubin*, Ann Rev L Soc Sci 2011, 17, 22: “If resources were no constraint, researchers should be able to articulate how one might design an experiment to study the question of interest.”; provokantes Wortspiel bei *Angrist/Pischke*, Econometrics 2009, 5: “If you can’t devise an experiment that answers your question in a world where anything goes, then the odds of generating useful results with a modest budget and nonexperimental survey data seem pretty slim. […] questions that cannot be answered by any experiment are FUQs: fundamentally unidentified questions.”

[224] *Börsch-Supan/Köke*, Ger Econ Rev 2002, 295, 319 f. nennt etwa statistische Methoden wie *robustness checks*, *instrumentation*, *sensitivity analysis* und *structural factor analysis*; zu einigen anschaulich *Legewie*, KZfSS 2012, 123, 130 ff.

[225] Früher Versuch bei *Barker*, Am Psy’st 1965, 1 (“ecological psychology”); in der Ökonomik sammelt *John List* seit 2004 Feldexperimente unter www.fieldexperiments.com.

[226] Qualitativ bspw. bei *Baumann/Reiss*, ZGR 1989, 157; *Lutter*, Letter of Intent 1998; quantitativ bei *Talley/O’Kane*, JITE 2012, 181.

[227] *Bortz/Döring*, Forschungsmethoden 2006, 149 ff.; *Diekmann*, Sozialforschung 2012, 576 ff.; *Heise*, Pepp L Rev 1999, 807, 825 ff.

[228] So *Röhl*, Rechtssoziologie 1987, 110 ff., abgrenzend zur qualitativen „Inhaltsanalyse“.

[229] Einerseits *Sedlmeier/Renkewitz*, Forschungsmethoden 2008, 121: „In der Psychologie gibt es im Grunde nur zwei Methoden, Daten zu erhalten: Befragen und Beobachten. […] bei Beobachtungsstudien muss man sich entscheiden […] ob man in das Geschehen eingreifen möchte oder nicht.“; andererseits *Hussy/Schreier/Echterhoff*, Forschungsmethoden 2010, 135:

selten berücksichtigte Vignettenmethode (c). Im Anschluss werfe ich die Frage nach der Konsistenz der so gewonnenen Ergebnisse auf (3.), bevor ich die thematische Klammer der Vorstandsvergütung schließe (V.).

a) Feldstudien (Beispiel: Gardner, Van Dyne & Pierce 2004)

Feldstudien sind nur dadurch definiert, dass sie in der „natürlichen Umgebung" der Untersuchungsteilnehmer ohne strukturelle Eingriffe der Untersuchungsleiter stattfinden.[230] Diese weite Definition umfasst auch Korrelationsstudien, daher bezeichne ich hier nur *primär*statistische Studien als Feldstudien – also solche, deren Urheber die benötigten Daten selbst aufzeichnet und folglich Einfluss darauf hat, welche Arten und Menge von Daten er erhält.

In einer 2004 veröffentlichten Feldstudie untersuchten drei Betriebswirte individuelle Daten von 91 Beschäftigten einer Baugesellschaft (von der Geschäftsführung abwärts) im nördlichen Mittelwesten der USA.[231] Die Datenerhebung erfolgte in drei Schritten, deren Rhythmus weniger theoretischen Erwägungen als vielmehr „den Erfordernissen des Betriebs geschuldet" war:[232] Zuerst wurde aus der Personalakte jedes Mitarbeiters sein im Vorjahr insgesamt verdientes Gehalt entnommen, drei Monate später wurden die Mitarbeiter über ihren organisationsbezogenen Selbstwert befragt,[233] und weitere neun Monate später wurden die vom Unternehmen regulär erstellten Personalbeurteilungen für die gesamten zwölf Monate seit Studienbeginn abgefordert.[234] Mit den so erhobenen Daten untersuchten die Autoren drei Hypothesen, deren im vorliegenden Zusammenhang relevanteste lautete, dass je höher das Vorjahresgehalt eines Mitarbeiters, desto höher seine diesjährige Arbeitsleistung sein würde. Dabei wurden in die Auswertung auch Daten über die Dauer der Betriebszugehörigkeit sowie etwaige Gehaltserhöhungen zwischen den Jahren aufgenommen.[235] Nach

---

„die vier klassischen experimentellen Untersuchungsformen" Labor-, Feld- und Quasiexperiment sowie Feldstudie; ähnl. auch *Shadish/Cook/Campbell*, Causal Inference 2002, 12 ff.: "Correlational Studies" sowie "Randomized", "Quasi" und "Natural Experiments".

[230] *Hussy/Schreier/Echterhoff*, Forschungsmethoden 2010, 136; *Bortz/Döring*, Forschungsmethoden 2006, 57.

[231] *Gardner/Van Dyne/Pierce*, J Occ Org Psy 2004, 307, 312: "Participants represented all job levels in the organization hierarchy including managerial, supervisory, professional and support staff."

[232] *Gardner/Van Dyne/Pierce*, J Occ Org Psy 2004, 307, 313.

[233] Mittels eines standardisierten 10-Punkte-Testbogens, der auch für fünf weitere Sprachen eingeführt wurde: *Kanning/Hill*, J Bus Media Psy 2012 (deutsche Fassung in Tab. 2).

[234] *Gardner/Van Dyne/Pierce*, J Occ Org Psy 2004, 307, 312: "Performance data were available for 62 of the study participants (due to turnover and missing records in the host organization)."

[235] *Gardner/Van Dyne/Pierce*, J Occ Org Psy 2004, 307, 314.

Berücksichtigung dieser Kontrollvariablen verblieb eine kleine, wenn auch statistisch signifikante Korrelation zwischen der Höhe des Vorjahresgehalts und der diesjährigen Personalbewertung:[236] Gut 4 % der diesjährigen Arbeitsleistung waren durch die vorjährige Vergütungshöhe zu erklären.[237] Diese Korrelation – so belegte die weitere Auswertung – kam durch den höheren organisationsbezogenen Selbstwert der gut bezahlten Mitarbeiter zustande,[238] damit legt die Studie nahe, dass ein höheres Vorjahresgehalt den Selbstwert steigert und dadurch die Arbeitsleistung positiv beeinflusst.

Offene Feldstudien wie die dargestellte wagen sich gewissermaßen einen Schritt weiter in den Bereich der Experimente vor als die Korrelationsstudie. Deshalb lassen sich Stärken und Schwächen der Feldstudie am leichtesten im Vergleich zur Korrelationsstudie erläutern:

Indem Gardner und Kollegen die meisten ihrer Daten aus den Personalakten bzw. dem regulären Personalbeurteilungsverfahren erhielten, konnten sie sicherstellen, authentische und unverfälschte Daten über Entlohnungshöhe und Arbeitsleistung zu erhalten, ähnlich wie in einer Korrelationsstudie. Dementsprechend konnten Gardner und Kollegen zwischen beiden Größen aber auch nur eine Korrelation feststellen und nichts über den Ursachenzusammenhang sagen. Arbeiteten Beschäftigte besser, *weil* sie im Vorjahr besser entlohnt wurden? Oder wurde ihre diesjährige Leistung von Vorgesetzten besser bewertet, weil schon im Vorjahr viel in den Mitarbeiter investiert worden war?[239] Oder wurden schlicht diejenigen Mitarbeiter im Vorjahr besser bezahlt *und* dieses Jahr besser bewertet, die von Vornherein besonders fähig / intelligent / gut ausgebildet waren? Für jede dieser drei (und viele weitere) Erklärungen lassen sich gute Argumente finden, und die Feldstudie kann zur Entscheidung kaum beitragen.[240] Darin ähnelt sie der Korrelationsstudie.

Sie *unterscheidet* sich von der Korrelationsstudie insoweit als sie aktiv ins Geschehen eingreift. Im zweiten Schritt der Datenerhebung erfragten Gardner und Kollegen von jedem Mitarbeiter persönliche Angaben, und lösten sich damit von der Beschränkung, nur vorhandene Daten verwenden zu können. Die Feldstudie kann daher auch Fragestellungen erschließen, die sich nicht aus bereits dokumentierten Zahlen beantworten lassen (etwa über das organisationsbezogene Selbstwertgefühl). Dafür gewärtigen Feld-

---

[236] *Gardner/Van Dyne/Pierce*, J Occ Org Psy 2004, 307, 315: "r = .21, p = .067".

[237] Genau 4,41 % beträgt das Bestimmtheitsmaß (oben § 2 A.IV.1.) aus der Korrelation von 0,21 (vorige Fn.).

[238] Mediationsanalyse in *Gardner/Van Dyne/Pierce*, J Occ Org Psy 2004, 307, 315.

[239] Diese Hypothese legen Studien zur Einsatzeskalation (unten § 4 C.IV.3.) nahe.

[240] *Hussy/Schreier/Echterhoff*, Forschungsmethoden 2010, 136: „die *Einbußen im Ausmaß der internen Validität* sind so erheblich, dass nur mit großer Zurückhaltung kausal interpretiert werden darf."

studien dieselben Gefahren, die schon eine Reihe früher Arbeitsplatzstudien in der US-amerikanischen Hawthorne-Manufaktur berühmt gemacht hatten.[241] Dort war die Produktivität unter verschiedenen Beleuchtungsbedingungen untersucht und (nach der legendären Überlieferung, die inzwischen widerlegt sein dürfte[242]) festgestellt worden, dass die Arbeiterinnen zwar produktiver arbeiteten, wenn ihr Arbeitsplatz besser beleuchtet wurde als zuvor, dass sie aber auch besser arbeiteten, wenn der Arbeitsplatz *schlechter* beleuchtet wurde als zuvor – nicht die Beleuchtung, sondern allein das Bewusstsein, beobachtet zu werden, war für die Produktivitätssteigerung verantwortlich. Die ungewöhnliche Aufmerksamkeit, die den Teilnehmern offener Feldstudien zuteil wird, kann also eine experimentelle Verunreinigung bilden und Schlussfolgerungen darüber verhindern, wie sich die Studienbedingungen (Beleuchtung, Bezahlung, o.ä.) auf die Produktivität auswirken.[243] Ähnliches könnte durchaus auch in der Studie von Gardner und Kollegen der Fall gewesen sein: Beschäftigte, die im Vorjahr gut bezahlt worden waren, könnten die zu Studienzwecken erfolgte Beobachtung als Bewährungsprobe empfunden und zum Anlass für höhere Arbeitsleistung genommen haben als andere Beschäftigte, die schon bisher nicht gut bezahlt wurden und deshalb nichts zu verlieren hatten. Der vermeintliche Effekt der Gehaltshöhe auf die Leistung wäre dann ein Artefakt, das ohne die Studie nicht entstanden wäre.

b) Befragungen (Beispiel: Pepper, Gore & Crossman 2013)

Eine der Feldstudie sehr ähnliche Studienform – und in der Tat ja sogar Bestandteil der eben erläuterten Feldstudie – ist die Befragung. Sie ist in der empirischen Sozialforschung „auf jeden Fall der meistbeschrittene Weg“[244] und unterscheidet sich (nach hiesigem Verständnis) vor allem dadurch von der Feldstudie, dass sie keine Daten aufzeichnet, die vom Studienteilnehmer losgelöst sind, sondern gerade dessen selbstberichtete Einstellungen und Wahrnehmungen.

---

[241] Die genauen Umstände dieser in den 1920er Jahren durchgeführten Studien sind umstritten (vgl. *Parsons*, Sci 1974, 922), was aber nur zur Legendenbildung beigetragen hat.

[242] Mehrere Neuauswertungen der Daten konnten den Hawthorne-Effekt nicht nachweisen: *Adair*, J Appl Psy 1984, 334; *Jones*, Am J Sociol 1992, 451; *Levitt/List*, Am Econ J Appl 2011, 224.

[243] Vgl. *Bortz/Döring*, Forschungsmethoden 2006, 504, die den Effekt aber i.R.d. *äußeren* Gültigkeit besprechen.

[244] *Diekmann*, Sozialforschung 2012, 434, und weiter: „Man hat das Interview gern als ‚Königsweg‘ der Sozialforschung bezeichnet. Das klingt schon deshalb gut, weil die Formulierung von Rene König (1972) stammt, einem der Begründer der modernen Sozialforschung im Deutschland der Nachkriegszeit.“

So befragten jüngst[245] zwei Betriebswirte und eine Organisationspsychologin etwa vier Siebtel aller Führungskräfte der 350 größten britischen Unternehmen.[246] Sie wollten die Effektivität und Effizienz von langfristigen Anreizprogrammen (*long term incentive plan*, auch LTI-Plan) überprüfen,[247] und fragten in diesem Zusammenhang die Unternehmensleiter auch nach ihrer Risikowahrnehmung und nach der Anreizwirkung der Vorstandsvergütung. Sie stellten fest, dass Unternehmensleiter einer systematischen Risikoüberschätzung unterliegen, weshalb die Erfolgsvergütung für das Unternehmen letztlich mehr kostet als sie durch Beeinflussung der Unternehmensleiter nützt.[248] Zudem bekundeten die Unternehmensleiter ihre Bereitschaft, für eine ideale Wunschposition auf die Hälfte ihres aktuellen Einkommens zu verzichten,[249] woraus man schließen könnte, dass der hohe verzichtbare Teil der Vergütung keine Anreizwirkung entfaltet, sondern lediglich Opportunitätskosten kompensiert. Die Autoren der Studie deuteten ihre Ergebnisse so, dass die Vergütungshöhe nur in einem mittleren Wertebereich die Leistungsbereitschaft positiv beeinflusse, während besonders niedrige Vergütungen als Ausdruck mangelnder Wertschätzung erscheinen und die Leistungsbereitschaft vermindern, besonders hohe Vergütungen dagegen die innere Motivation zerstören (*crowding-out*, oben bei Fn. 149) und damit ebenfalls kontraproduktiv wirken.[250] Bis zu welcher Obergrenze die Vergütung noch leistungssteigernd wirkt, kann die Studie freilich allenfalls aufgrund anekdotischer Evidenz mutmaßen.[251]

---

[245] Ältere Befragungen zur (US-amerikanischen) Vorstandsvergütung z.B. *Tosi/Gomez-Mejia*, Admin Sci Q 1989, 169; *Zajac*, Strat Mgmt J 1990, 217 (“this study is interested explicitly in CEO compensation as a motivational tool that can enhance firm performance.”, 221).

[246] *Pepper/Gore/Crossman*, Hum Res Mgmt J 2013, 36, 40: “905 survey participants were selected randomly from a sample frame of 1,563 individuals, being the number of senior executives working for FTSE 350 companies based on a detailed examination of each company’s website, most recent annual report and other public data.”; der Index *Financial Times and (London) Stock Exchange* (FTSE) 350 umfasst die nach ihrer Marktkapitalisierung größten Unternehmen mit Erstnotierung an der Londoner Börse.

[247] *Pepper/Gore/Crossman*, Hum Res Mgmt J 2013, 36 f.

[248] *Pepper/Gore/Crossman*, Hum Res Mgmt J 2013, 36, 48: “the value of a long-term incentive, as mentally accounted for by a senior executive, is likely to be less than the amount which the company providing the incentive has to account for”; mit Blick aufs deutsche Recht näher *Hamann*, BRJ 2010, 27, 29 f.

[249] *Pepper/Gore/Crossman*, Hum Res Mgmt J 2013, 36, 46: “The results varied between a minimum discount of 0.00 and a maximum of 0.92, with a mean discount score of 0.48, a median score of 0.50 and a standard deviation of 0.24.”

[250] *Pepper/Gore/Crossman*, Hum Res Mgmt J 2013, 36, 48: “varied at the top end because of weak crowding out, and at the bottom end by demoralisation costs, giving an angled, inverted ‘S’ shape. Thus, [only] in the middle range of the curve, effort increases monotonically with additional reward”.

Befragungsstudien sind sowohl in ihrer inneren als auch ihrer äußeren Gültigkeit stark beschränkt. Indem sie sich auf Eigenauskünfte des untersuchten Personenkreises stützen, genießen sie zwar die größtmögliche äußere Gültigkeit, was die *geäußerten* Ansichten und Meinungen dieses Personenkreises angeht. Ob diese Personen allerdings die von ihnen *tatsächlich* erlebten Zustände – oder auch nur ihre tatsächlichen *Ansichten und Meinungen* darüber – wiedergeben (wollen und können), ist nie sicher. Erstens besteht die Gefahr, dass die Befragten „sich gegenüber dem Interviewer positiv darzustellen hoffen"[252] und ihre Selbstdarstellung oder den Studienausgang bewusst beeinflussen – und zwar umso mehr, je transparenter der Befragungszweck für die Befragten und je größer ihr Interesse an einem bestimmten Ausgang der Studie ist. In der Studie von Pepper und Kollegen könnte den befragten Führungskräften durchaus daran gelegen gewesen sein, sich als weniger vergütungsorientiert darzustellen[253] als dies tatsächlich der Fall ist. Zweitens lässt sich nie abschätzen, ob die Befragten nicht unbewusst beeinflusst wurden, etwa durch die Umgebung, in der sie die Fragen beantworteten, durch selbst auferlegte Zeitrestriktionen, durch mangelndes Verständnis einzelner Fragen, usw. Indem eine Befragungsstudie die Datenerhebung nahezu vollständig aus der Hand gibt, besteht stets die Gefahr, dass Fragen anders wahrgenommen und beantwortet werden als vom Fragenersteller beabsichtigt. Dem lässt sich dadurch vorbeugen, dass Kontrollfragen eingefügt oder mehrere Antworten zu Antwortskalen zusammengefasst werden, allerdings nimmt mit der Länge des Fragebogens wiederum die Teilnahmebereitschaft ab (gerade bei stark beanspruchten Studienteilnehmern wie Unternehmensleitern). Nicht umsonst werden Pepper und Kollegen ihren Fragebogen auf insgesamt 15 Fragen beschränkt haben.[254]

Sieht man von den möglichen Antwortverzerrungen ab, hängt die Übertragbarkeit von Befragungsergebnissen vor allem davon ab, welcher Teil der Befragten überhaupt antwortet. Bei Befragungsstudien allgemein und bei der Befragung von Führungskräften mit hoher Arbeitsauslastung ganz besonders ist die Rücklaufquote (*response rate*) normalerweise eher gering. Die Studie von Pepper und Kollegen ist dafür ein gutes Beispiel: Von

---

[251] *Pepper/Gore/Crossman*, Hum Res Mgmt J 2013, 36, 43: "One company chairman, commenting specifically on the US market, said: 'I do not believe, nor have I ever observed, that $100 million motivates people more than $10 million, indeed more than $1 million'."

[252] *Röhl*, Rechtssoziologie 1987, 109 („Social-Desirability-Effekt").

[253] Vgl. *Pepper/Gore/Crossman*, Hum Res Mgmt J 2013, 36, 41: "the prevailing view was that […] only a small number of executives are primarily motivated by potential monetary gain, perhaps no more than 10 per cent or 20 per cent according to one HR director."; vgl. auch das Zitat oben Fn. 150.

[254] *Pepper/Gore/Crossman*, Hum Res Mgmt J 2013, 36, 40.

905 Befragten meldeten sich überhaupt nur 140 zurück, von denen 65 die Teilnahme ablehnten.[255] Mithin antworteten effektiv nur 75 der 905 Befragten, was mit 8,3 % eine durchaus übliche Rücklaufquote im einstelligen Prozentbereich ergibt.[256] Das schränkt nicht zwangsläufig die Übertragbarkeit der Ergebnisse ein, solange keine systematischen Unterschiede zwischen den Antwortenden und den Nichtantwortenden bestehen. Mindestens zwei solcher Unterschiede werden in der Studie allerdings explizit angesprochen: Von den 65 Befragten, die die Teilnahme ausdrücklich ablehnten, wiesen einige auf entgegenstehende Unternehmensrichtlinien hin, andere auf Arbeitsüberlastung.[257] Das bedeutet, dass die Ergebnisse wohl nur von Unternehmensleitern stammen, die ausreichend Muße, Kooperationsbereitschaft und unternehmensinterne Freiheit haben, an wissenschaftlichen Studien teilzunehmen. Dabei ist für die 765 stillschweigenden Ablehnungen noch nicht einmal ein Grund bekannt, was die Abschätzung der Übertragbarkeit noch zusätzlich erschwert. Letztlich bedarf es deshalb nach der Übertragungslehre (oben bei Fn. 24) einer ausführlichen Argumentation, warum die Ergebnisse dennoch auf Vorstandsmitglieder im Allgemeinen übertragbar sein sollten.

Die innere Gültigkeit von Befragungsstudien ist regelmäßig minimal, da die befragten Personen etwaige Ursachenzusammenhänge auch nicht zuverlässiger identifizieren dürften als die Forscher selbst aus ihrer Lebenserfahrung heraus (dazu oben § 2 A.I.1.). Daher dienen Befragungsstudien eher der Gewinnung als der Überprüfung von Ursachenhypothesen und sind recht selten in der quantitativen, aber deutlich öfter im Rahmen der qualitativen Forschung (dazu § 1 D.I.) anzutreffen; auch der erste, hier nicht näher dargestellte Teil der Studie von Pepper und Kollegen war rein qualitativ.[258] Eher für die quantitative Forschung geeignet und gedacht sind Panelbefragungen,[259] also wiederholte systematische Querschnittserhebungen (vgl. oben § 2 bei Fn. 78).

---

[255] *Pepper/Gore/Crossman*, Hum Res Mgmt J 2013, 36, 40.

[256] Ähnlich die an anderer Stelle dieser Arbeit referierten Befragungsstudien: 7,25 % bei *Hermes*, Haftungsbeschränkung 2006, 52 (oben § 1 bei Fn. 269), 7,8 % bei *Trottier/Gordon*, SSRN 2011, 11 (oben bei Fn. 202); für die Management-Forschung dagegen berichten *Cycyota/Harrison*, Org Res Meth 2006, 133 (Metastudie über 231 Primärstudien 1992–2003) eine Rücklaufquote von 32 % im Median, mit Interquartilsabstand 20–46 %.

[257] *Pepper/Gore/Crossman*, Hum Res Mgmt J 2013, 36, 40: "65 declined to participate saying that it was against company policy to do so or they were too busy".

[258] *Pepper/Gore/Crossman*, Hum Res Mgmt J 2013, 36, 39: "based around a programme of semi-structured interviews, […] Stage 1 comprised a qualitative study of 15 senior executives […] The participants were identified via the first researcher's professional contacts, a form of convenience sampling."

[259] Etwa die in *KfW*, Publikationen 2012 oder (für die Schweiz) in *ETH-KOF*, Strukturumfragen 2012 nachgewiesenen.

c) Vignettenstudien (Beispiel: Weibel, Rost & Osterloh 2007)

Vignettenstudien sind gewissermaßen ein Hybrid aus der Befragung und experimentellen Ansätzen, daher werden sie auch als Vignettenexperimente[260] oder Befragungsexperimente (*survey experiments*)[261] bezeichnet. Die Methode geht auf eine 1951 veröffentlichte Dissertation zurück, hat aber „von Ausnahmen abgesehen, als Forschungsmethode relativ wenig Beachtung gefunden."[262] Womöglich beruht der vermeintliche Achtungsmangel aber auch darauf, dass sich Vignettenforschung in zahlreichen Disziplinen („neben den Sozialwissenschaften etwa auch in den Gesundheitswissenschaften, der Rechtswissenschaft, der Psychologie und der Marktforschung."[263]) unter ganz verschiedenen Synonymen versteckt, von Szenariostudien (*scenario-based research*) in der Ethikforschung[264] über faktorielle Surveys (*factorial surveys*) in der Soziologie[265] und experimenteller Strategieerfassung (*policy-capturing*) in der Betriebswirtschaftslehre[266] bis hin zu einer Unmenge anderer Begriffe für diese und ähnliche Methoden.[267] Man könnte alle unter den Oberbegriff „Dekompositionsmethoden" fassen, weil sie Entscheidungen danach aufzugliedern versuchen, durch welche Entscheidungsfaktoren sie verursacht werden.[268] Eine solche Studie, die sich unlängst dem Entscheidungsfaktor „Vorstandsvergütung" widmete, stammt von drei Betriebswirtinnen aus der Schweiz.[269]

---

[260] *Kunz/Linder*, ZPU 2011, 211; *Weibel/Rost/Osterloh*, zfbf 2007, 1029, 1038 m.w.N.

[261] Bspw. im Sonderheft der Zeitschrift "Methodology", vgl. Einl. von *Schlueter/Schmidt*, Meth 2010, 93; als „Befragung zu *fiktiven Fällen*" fasst sie *Röhl*, Rechtssoziologie 1987, 108.

[262] *Diekmann*, Sozialforschung 2012, 348 m.Verw. auf *Beck/Opp*, KZfSS 2001, 283 f.; *Atzmüller/Steiner*, Meth 2010, 128: "too infrequently used research method" m.w.N. auf 129; *Wallander*, Soc Sci Res 2009, 505, 506: "most sociologists have never heard of it."; *Kunz/Linder*, ZPU 2011, 211, 220 zur „bisher eher geringen Nutzung von Vignetten-Experimenten in der betriebswirtschaftlichen Forschung"; dagegen sehen *Auspurg/Hinz u.a.*, soFid Methoden 2009, 2/23, 24 einen „Kontrast" zwischen der „häufigen Anwendung" und der „geringen Erforschung des Verfahrens"; ähnl. *Auspurg/Hinz/Liebig*, MDA 2009, 59, 87.

[263] *Auspurg/Hinz/Liebig*, MDA 2009, 59, 60.

[264] Bspw. *Weber*, Bus Ethics Q 1992, 137 (Forschungsbericht über 26 Studien).

[265] Grundlegend *Rossi/Nock*, Factorial Survey 1982; *Wallander*, Soc Sci Res 2009, 505 (Forschungsbericht über 106 Studien 1982–2006); ausf. *Jasso*, Sociol Meth Res 2006, 334.

[266] *Zedeck*, Org Behav Hum Perf 1977, 47; *Karren/Barringer*, Org Res Meth 2002, 337 f.

[267] *Aiman-Smith/Scullen/Barr*, Org Res Meth 2002, 388, 388 f.: "In marketing it is called *conjoint analysis*, *stated preference models*, or *trade-off analysis*. Researchers on problem-solving with computer and information systems use the term *protocol analysis*. In environmental and social policy research, the terms *contingent preference*, *contingent choice*, or *analytical hierarchical process models* are often used. Strategy and OB/HR researchers generally use the term *policy-capturing*. […] Despite the differences in terminology, actual differences in the techniques are minimal."

[268] *Karren/Barringer*, Org Res Meth 2002, 337, 355 m.w.N.

Die Autorinnen gehen von der Beobachtung aus (vgl. schon oben bei und in Fn. 149), dass „extrinsische und intrinsische Motivation nicht voneinander unabhängig [sind], d.h. die beiden Motivationsarten wirken nicht additiv.“[270] Sie vermuten, dass erfolgsabhängige Vergütung zwar einerseits die äußere (extrinsische) Motivation stärkt, aber andererseits die innere (intrinsische) Motivation schwächt, so dass die „relative Stärke beider gegenläufiger Effekte“ darüber entscheidet, ob erfolgsabhängige Vergütung insgesamt die Leistung steigern kann.[271] Um diese Vermutung zu überprüfen, legten die Autorinnen 149 Berufstätigen (davon zwei Drittel in leitender Funktion)[272] Situationsbeschreibungen aus dem Berufsleben vor und veränderten innerhalb dieser Beschreibungen zehn verschiedene Situationsfaktoren, die jeweils mehrere Ausprägungen annehmen konnten. Der wichtigste Situationsfaktor „variabler Lohn“ erhielt drei Ausprägungen:

„Ihr eigener Lohn variiert nicht. Sie haben ein fixes Gehalt.“ bzw.

„Ihr eigener Lohn variiert nur gering. Er besteht zu 95 % aus einem Fixanteil und zu 5 % aus einem leistungsabhängigen Bonus.“ bzw.

„Ihr eigener Lohn ist sehr variabel. Er besteht zu 50 % aus einem Fixanteil und zu 50 % aus einem leistungsabhängigen Bonus.“[273]

Die anderen neun Situationsfaktoren waren in der Veröffentlichung nicht erwähnt, lauteten aber nach Auskunft der Zweitautorin:[274]

„selbstbestimmtes Arbeiten“ – „Tätigkeit bestätigt Ihre Kompetenzen“ – „Kollegen mögen und schätzen Sie“ – „Projektvorgesetzter mag Sie“ – „Möglichkeiten zur Mitbestimmung“ – „mehr Leistung = Weiterkommen“ – „hohe Transparenz und Objektivität der

---

[269] *Weibel/Rost/Osterloh*, zfbf 2007, 1029; auf Englisch erneut veröffentlicht in *Weibel/Rost/Osterloh*, J Publ Admin Res Theory 2010, 387, 397 ff.

[270] *Weibel/Rost/Osterloh*, zfbf 2007, 1029, 1033 und weiter: „Dieser Effekt wird als ‚hidden costs of reward‘ oder als Verdrängungseffekt (‚crowding-out-effect‘) bezeichnet.“; vgl. auch *Stenzel*, Vorstandsvergütung 2012, 224 f., der noch aus einer Vorabversion von *Weibel/Rost/Osterloh*, J Publ Admin Res Theory 2010, 387 zitiert.

[271] *Weibel/Rost/Osterloh*, zfbf 2007, 1029, 1035; vgl. auch *Schüller*, Vorstandsvergütung 2002, 56; *Weibel/Rost/Osterloh*, J Publ Admin Res Theory 2010, 387, 397 f.: “On the one hand, performance-contingent rewards subdue the internalized meaning of the work itself, that is, lowers [sic] the power of intrinsic motivation […] On the other hand, pay for performance strengthens the meaning of external rewards, that is, […] causes a price effect on motivation and performance. The overall effect thus is dependent on the relative strength of the two unobservable contradicting effects.”

[272] Näher *Weibel/Rost/Osterloh*, zfbf 2007, 1029, 1035: „Die [.] Erhebung fand im Jahre 2006 in einer berufsbegleitenden Fachhochschule statt. […] Probanden verfügen durchschnittlich über neun Jahre Berufserfahrung im Dienstleistungsbereich […] 67 % der Teilnehmenden haben derzeit im Unternehmen eine Leitungsfunktion inne.“

[273] *Weibel/Rost/Osterloh*, zfbf 2007, 1029, 1042.

[274] Professorin Katja *Rost* per E-Mail vom 24.11.2012.

Beurteilung und Belohnung" – „Projektvorgesetzter behandelt Sie respektvoll" – „Ihr Unternehmen vermeidet Arbeitsplatzabbau".

Jeder dieser neun Situationsfaktoren hatte zwei Ausprägungen (ja / nein), für alle zehn Situationsfaktoren waren also 3 x 2^9 = 1.536 verschiedene Kombinationen denkbar. Von diesen 1.536 möglichen Situationsbeschreibungen legten die Autorinnen jedem Teilnehmer drei vor, so dass insgesamt 447 Situationsbeschreibungen beurteilt wurden.[275] (Wie die Stichprobe ausgewählt wurde, erwähnt die Studie leider nicht.) Jede Situation sollte daraufhin beurteilt werden, wieviel freiwillige Mehrarbeit der Teilnehmer in dieser Situation zu leisten bereit wäre,[276] und aus welcher Motivation heraus.[277] Die so gesammelten Daten belegten in der statistischen Auswertung einen positiven Gesamteffekt der variablen Vergütung auf den Arbeitseinsatz.[278] Zugleich zeigte die Auswertung aber auch,[279] dass dieser positive Gesamteffekt nur dann zustande kam, wenn die extrinsische Motivation von vornherein überwog.[280] Die Autorinnen resümmierten daher:

„Die Ankündigung variabler Löhne führt stets zu einer Verdrängung der intrinsischen Motivation. Dieser Verdrängungseffekt wird durch den relativ höheren Preiseffekt [nur dann] überkompensiert, so dass ein positiver Totaleffekt von variablen Löhnen auf die (intendierte) Arbeitsleistung resultiert, [… wenn] bei einer Tätigkeit die extrinsische Motivation [überwiegt]"[281]

Vignettenstudien wie die vorliegende sind stets hybrid: Einerseits sind sie *Befragungen*, erzeugen also hohe äußere Gültigkeit. Diese beschränkt sich zwar auf Absichtserklärungen – und „Einstellung und Verhalten stimmen mitunter nicht überein."[282] Allerdings sind Vignetten deutlich konkreter formuliert und weniger interpretationsbedürftig als die unscharfen und ab-

---

[275] Missverständlich insoweit *Weibel/Rost/Osterloh*, zfbf 2007, 1029, 1041: „Wir erfragen das komplette Vignettenuniversum über mehrere Probanden." (ebenso *Weibel/Rost/Osterloh*, J Publ Admin Res Theory 2010, 387, 400).

[276] Anzugeben war die beabsichtigte wöchentliche Mehrarbeit auf einer Skala mit den fünf Antwortmöglichkeiten „> 5 h", „≤ 3-4 h", „≤ 2 h", „≤ 1 h" und „keine Stunden" (*Weibel/Rost/Osterloh*, zfbf 2007, 1029, 1042).

[277] Zum Einsatz kam ein Fragebogen, der in anderen Zusammenhängen bereits als Messinstrument für die unterschiedlichen Motivationsarten validiert wurde (*Weibel/Rost/Osterloh*, zfbf 2007, 1029, 1044 m.w.N.).

[278] *Weibel/Rost/Osterloh*, zfbf 2007, 1029, 1047: „Modell 2 zeigt, dass variable Löhne die Intention unserer Befragten für zusätzliches Arbeitsengagement erhöhen (β = 0,24***)."

[279] *Weibel/Rost/Osterloh*, zfbf 2007, 1029, 1047: „Modell 4 berücksichtigt die Interaktion zwischen Motivation und variablem Lohn. […] Der [.] Totaleffekt von variablen Löhnen auf zusätzliche Arbeitsleistung ist nicht mehr signifikant."

[280] *Weibel/Rost/Osterloh*, zfbf 2007, 1029, 1048: „wie zum Bespiel [sic] Arbeiten am Fliessband"; intrinsische Motivation hingegen überwiege „für die Ausführung komplexer, kreativer Aufgaben" (ebd. 1037); vgl. auch § 4 Fn. 397.

[281] *Weibel/Rost/Osterloh*, zfbf 2007, 1029, 1048.

strakten Fragen der meisten Befragungsstudien.[283] Indem sie „realitätsnahe Alltagssituationen […] durch Geschichten (Story-Telling) [er]schaffen, innerhalb derer die variierten Merkmale versteckt sind“, sprechen Vignetten die Lebenserfahrung der Teilnehmer an und „erleichtern den Befragten […] lebensnahe Rolleninterpretationen.“[284] Zugleich verhindert das „Verstecken“ der variierten Merkmale, dass der Studienzweck erfasst und gezielt darauf reagiert wird.[285] Zwar „können die Resultate im Vignettenexperiment – wie in allen Befragungen – durch die Angabe sozial erwünschter Verhaltensweisen beeinträchtigt werden.“[286] Allerdings sind Vignetten im Vergleich zu anderen Befragungen „weniger direkt, sodass Verzerrungen durch ‚soziale Wünschbarkeit‘ möglicherweise eine geringere Rolle spielen.“[287]

Andererseits sind Vignettenstudien auch „richtige Experimente“,[288] denn sie weisen ihren Teilnehmern im Längsschnitt (*within subjects*) einige von vielen systematisch variierten Situationsbeschreibungen zufällig zu (vgl. noch Fn. 298). Das ermöglicht Ursachenschlüsse, weshalb die innere Gültigkeit von Erkenntnissen aus Vignettenstudien „als sehr hoch eingeschätzt“ wird.[289] Dabei ermöglichen sie mehr als andere Experimente die Berücksichtigung „weicher“ Kontextfaktoren, denn Vignetten „zeichnen sich – wie tatsächliche soziale Situationen – durch Mehrdimensionalität aus.“[290] Das heißt, dass sie „die Konstruktion sehr umfassender Situationen

---

[282] *Weibel/Rost/Osterloh*, zfbf 2007, 1029, 1039; *Weibel/Rost/Osterloh*, J Publ Admin Res Theory 2010, 387, 401, 405: “the vignette study has the drawback not to observe real behavior but only behavioral intentions.”; auch der Methodenvergleich von *Groß/Börensen* in: Kriwy/Gross, Klein aber Fein 2009, 149, 172 ergab, „dass die Messung von tatsächlichem Verhalten mit dem faktoriellen Survey offenbar kaum möglich ist.“; *Finch*, Sociol 1987, 105, 113: “It is in this area […] that I see the biggest danger of the misuse of vignettes […] since many surveys do in fact seek to predict actions on the basis of responses given.”; vgl. oben nach Fn. 251.

[283] *Steiner/Atzmüller*, KZfSS 2006, 117, 118; *Alexander/Becker*, Publ Opin Q 1978, 93 f., 103: “The use of vignettes helps to standardize the social stimulus across respondents”; *Finch*, Sociol 1987, 105, 106.

[284] *Weibel/Rost/Osterloh*, zfbf 2007, 1029, 1038 f.; *Alexander/Becker*, Publ Opin Q 1978, 93, 103: “The use of vignettes […] makes the decision-making situation more real.”; *Finch*, Sociol 1987, 105, 111.

[285] *Wallander*, Soc Sci Res 2009, 505, 506: “respondents in factorial survey studies are most probably not fully attentive to the manipulation”.

[286] *Weibel/Rost/Osterloh*, zfbf 2007, 1029, 1039; *Groß/Börensen* in: Kriwy/Gross, Klein aber Fein 2009, 149, 173; *Kunz/Linder*, ZPU 2011, 211, 218 f.

[287] *Diekmann*, Sozialforschung 2012, 348; *Alexander/Becker*, Publ Opin Q 1978, 93, 95: “the respondent is not as likely to consciously bias his report in the direction of impression-management (social approval of the interviewer)”.

[288] *Kunz/Linder*, ZPU 2011, 211, 215.

[289] *Kunz/Linder*, ZPU 2011, 211, 214.

[290] *Weibel/Rost/Osterloh*, zfbf 2007, 1029, 1039; *Beck/Opp*, KZfSS 2001, 283, 304.

mit einer relativ hohen Anzahl an Situationsvariablen“ ermöglichen, anhand derer es gelingt, „die Einflussgewichte komplexer Ursachenkonstellationen zu ermitteln.“[291] Das gilt insbesondere für die „Messung von Normen“,[292] in deren Rahmen

> „die Befragten bei der Vignettenbewertung einerseits ihre Wertorientierungen nicht explizit nennen müssen und andererseits durch die Vielzahl an gleichzeitig präsentierten Vignettenmerkmalen die normative Orientierung erschwert wird.“[293]

Zudem wird es bisweilen als Vorteil der Vignettenstudie angesehen, dass sie ihre Teilnehmer auch Situationen aussetzen kann,

> „die realiter eher selten auftreten oder noch gar nicht existieren. Hierdurch wird im Vergleich zu Feldstudien und -experimenten die Auslotung theoretischer Randbereiche erleichtert“.[294]

Insgesamt gilt die Vignettenstudie daher als vielversprechende Kombination aus der äußeren Gültigkeit einer Befragung mit der inneren Gültigkeit eines Experiments,[295] wenngleich auch diese Kombination zahlreiche methodische Abwägungsschritte erfordert. Insbesondere darf die Anzahl der variierten Dimensionen nicht zu groß sein, um Teilnehmer nicht kognitiv zu überfordern, aber auch nicht zu klein, da „bei sehr inhaltsleeren Vignetten der experimentelle Vorteil einer hohen Kontrolle über das Untersuchungsdesign verloren“ gehen kann (vgl. oben bei Fn. 141).[296] Auch darf die Anzahl der Vignetten, die jedem Teilnehmer vorgelegt werden, nicht so groß sein, dass die Teilnehmer „merken, welche Merkmale variieren“ und „ihre Antworten im Hinblick auf diese ‚politisch korrekt‘ abstimmen“ können[297] – aber auch nicht so klein, dass die statistische Auswertbarkeit der Antworten beeinträchtigt wird.[298] Die genaue Ausgestaltung der Vignettenstudie entscheidet also über ihre methodische Nähe zu Befragungs- oder

---

[291] *Kunz/Linder*, ZPU 2011, 211, 214 (erstes Zitat); *Diekmann*, Sozialforschung 2012, 348 (zweites Zitat); *Auspurg/Hinz/Liebig*, MDA 2009, 59, 62; *Alexander/Becker*, Publ Opin Q 1978, 93, 95: “the vignette allows for a rather precise estimate of the effects of changes in *combinations* of variables”.

[292] So im Titel von *Beck/Opp*, KZfSS 2001, 283, 286; *Weibel/Rost/Osterloh*, zfbf 2007, 1029, 1038; *Diekmann*, Sozialforschung 2012, 347; ebenso schon *Finch*, Sociol 1987, 105.

[293] *Steiner/Atzmüller*, KZfSS 2006, 117, 134; zust. zit. von *Rabl*, ZfB 2012, 5, 16.

[294] *Kunz/Linder*, ZPU 2011, 211, 214; krit. zur Glaubwürdigkeit solcher Vignetten *Finch*, Sociol 1987, 105, 111; *Auspurg/Hinz/Liebig*, MDA 2009, 59, 88 f.; vgl. auch *Wallander*, Soc Sci Res 2009, 505, 507 m.w.N.

[295] *Schlueter/Schmidt*, Meth 2010, 93: “a flexible and powerful means to combine the internal validity of experiments with the external validity of surveys.”; *Atzmüller/Steiner*, Meth 2010, 128: “Vignette studies combine ideas from classical experiments and survey methodology to counterbalance each approach’s weakness.”; *Diekmann*, Sozialforschung 2012, 346.

[296] *Auspurg/Hinz u.a.*, soFid Methoden 2009, 2/23, 25.

[297] *Auspurg/Hinz u.a.*, soFid Methoden 2009, 2/23, 26.

Experimentalstudien und wirkt sich dementsprechend auf die Gültigkeiten der dadurch gewonnenen Erkenntnisse aus.

d) Quasi-Experimente (Beispiel: Bayer & Burhop 2008)

Ähnlich wie Vignettenstudien stehen auch Quasi-Experimente gewissermaßen an dem Punkt, an dem die beobachtende Methodik in die experimentelle übergeht. Dementsprechend gibt es zwei mögliche Definitionsansätze: Aus Richtung der beobachtenden Methodik betrachtet, sind Quasi-Experimente reguläre Korrelationsstudien mit einem systematischen Bruch im Datensatz. Aus Richtung der experimentellen Methodik betrachtet, sind Quasi-Experimente reguläre (Labor- oder Feld-)Experimente ohne die zufällige Zuweisung zu Versuchsbedingungen.[299] Das bedeutet, dass Quasi-Experimente zwar keine Zufallszuweisung verwenden, dass sie aber einen Bruch in den Daten ausnutzen, durch den sie der Eindeutigkeit einer Zufallszuweisung recht nahe kommen.[300] Eine an früherer Stelle dieser Arbeit bereits dargestellte Variante des Quasi-Experiments ist die Ereignisstudie (§ 1 G.II.), eine andere besonders einfallsreiche Variante ist die Auswertung von Schwellenwertsprüngen (*regression discontinuity design*).[301] Wie jene, vergleicht auch eine weitere Form des Quasi-Experiments – das *natürliche Experiment* – Daten vor und nach dem Auftreten einer äußeren Ursache, um aus dem Vorher-Nachher-Vergleich die Auswirkungen dieser Ursache abzulesen.

In einer 2008 als Vorstudie und 2009 in erweiterter Fassung veröffentlichten Studie nutzten zwei deutsche Volkswirte die Aktienrechtsreform von 1884 als natürliches Experiment, um die Anreizwirkung der Vorstandsvergütung zu hinterfragen.[302] Diese Studie fällt ein wenig aus dem Rahmen der vorliegenden Fragestellung, denn die Studienautoren untersuchten nicht die Anreizwirkungen der vorab vereinbarten Vergütung[303] sondern nur nachträglich gewährte Bonuszahlungen, deren Anreizwirkung bekanntlich

---

[298] *Atzmüller/Steiner*, Meth 2010, 128, 129: "In between-subjects designs each respondent judges only one single vignette, but they are quite rare since serious measurement problems arise, due to different vignette contexts."; dagegen *Auspurg/Hinz/Liebig*, MDA 2009, 59, 62 Fn. 3: „Es gibt durchaus Argumente, […] nur eine einzige Vignette zu präsentieren: Die Effekte sozialer Erwünschtheit sowie […] Lerneffekte werden vermindert."

[299] *Diekmann*, Sozialforschung 2012, 356, 330: „Quasi-Experimente sind in der Hauptsache Experimente ohne Randomisierung."; z.B. *Braun/Eidenmüller u.a.*, ZHR 2013, 131, 138.

[300] *Ho/Rubin*, Ann Rev L Soc Sci 2011, 17, 20: "The unifying feature of this movement is the attempt to hew as closely as possible to an experiment."; ebenso *Gangl/DiPrete*, KZfSS Sonderheft 2004, 396, 397; ausf. *Legewie*, KZfSS 2012, 123, 130 ff.

[301] Dazu *Legewie*, KZfSS 2012, 123, 138 ff., 140: „ein Design, dass [sic] nur in sehr speziellen Situationen zur Anwendung kommen kann […] aber in der Regel sehr verlässliche Schätzer hervor"bringt; gesellschaftsrechtliches Bsp. bei *Listokin*, Am L Econ Rev 2008, 159.

[302] *Bayer/Burhop*, sbr 2008, 378; *Bayer/Burhop*, Expl Econ Hist 2009, 464.

differenziert zu behandeln ist.[304] Ein für die vorliegende Frage besser passendes Quasi-Experiments hätte Vorstände vergleichen müssen, die allein aufgrund äußerer Umstände unterschiedliche Vergütungen erhalten, allerdings war eine solche Studie nicht auffindbar.[305] Die Arbeit von Bayer und Burhop ist aber für die Frage, wie Vergütungsstruktur und Anreizwirkung zusammenhängen, ebenfalls relevant und jedenfalls sehr instruktiv, was das methodische Vorgehen angeht. Da es hier in erster Linie um eine Illustration empirischer Methoden geht, erscheint mir die Studie trotz ihrer leicht divergierenden Forschungsfrage als ausgezeichnetes Anschauungsmaterial.

Die Studie setzt bei der rechtsökonomischen These an, dass die dem Vorstand gewährte Vergütung nicht nur Leistungsanreize setzen, sondern auch die Gefahr kompensieren muss, dass der Vorstand sich aus dem Gesellschaftsvermögen bereichert.[306] Daher vermuteten die Studienautoren, dass je vorstandsfreundlicher das zwingende Aktienrecht ausgestaltet ist, desto stärker das Interesse der Aktionäre daran sein müsse, die Vorstandsvergütung an den Unternehmenserfolg zu koppeln.[307] Diese Vermutung überprüften die Autoren anhand von Daten über die Vorstandsvergütung in 9 Banken sowie 37 Unternehmen des produzierenden Gewerbes im Zeitraum 1870–1911.[308] Mitten in diesem Zeitraum erfolgte die Aktienrechtsreform von 1884, die dem damals verbreiteten Gründungsbetrug unter anderem durch Verschärfung der Überwachungspflichten und der Haftung von Vorstand und Aufsichtsrat einen Riegel vorschieben sollte.[309] Da diese Art der Krisenregulierung den Rechtsrahmen abrupt änderte statt einen sanft gewachsenen Übergang nachzuvollziehen, sahen die Studienautoren darin

---

303 *Bayer/Burhop*, sbr 2008, 378, 388: “we do not have information about the fixed salary”; ebenso *Bayer/Burhop*, Expl Econ Hist 2009, 464, 470.

304 BGHSt 50, 331, 337: „Unter dem Gesichtspunkt einer Anreizwirkung für Dritte erscheint die Zuwendung einer freiwilligen Anerkennungsprämie […] denkbar, [… aber] entzieht sich generalisierender Betrachtung“; dazu bspw. *Ihrig/Wandt/Wittgens*, ZIP 2012, H. 40 Beil., 14 f.

305 Allenfalls gibt es ähnliche Studien wie die vorliegende – z.B. *Cuñat/Guadalupe*, J Bank Fin 2009, 495, die Deregulierungen im US-amerikanischen Finanzsektor 1994 und 1999 als natürliches Experiment für den Einfluss von Wettbewerb auf Höhe und Erfolgsabhängigkeit der Vorstandsgehälter nutzten.

306 Ausf. *Bayer/Burhop*, Expl Econ Hist 2009, 464, 465.

307 Ökonomisch formalisiert bei *Bayer/Burhop*, sbr 2008, 378, 382 ff. und (anders) bei *Bayer/Burhop*, Expl Econ Hist 2009, 464, 467.

308 *Bayer/Burhop*, Expl Econ Hist 2009, 464 f., 469 (471: “Our analysis focuses on the manufacturing sample, using banks primarily for comparison.”); *Bayer/Burhop*, sbr 2008, 378, 386 beschränkten sich auf die neun Banken.

309 *Hommelhoff* in: Schubert/Hommelhoff, Modernes AktR 1985, 53, 64: „Zentrales Anliegen des Reformgesetzes von 1884 war es, das Gründungsrecht der Aktiengesellschaft […] zu verbessern.“ und weiter 85 ff. zur Reform der Organisationsverfassung.

eine günstige Gelegenheit, dieselben Unternehmen zum fast gleichen Zeitpunkt einmal mit und einmal ohne die neuen anlegerschützenden Regelungen zu betrachten. Sie stellten fest, dass sich nach 1884 der als Bonuszahlung an den Vorstand ausgeschüttete Anteil des Unternehmensgewinns etwa halbierte.[310] Daraus schlossen Bayer und Burhop, dass vor der Reform ein erheblicher Anteil (eben ungefähr die Hälfte) der Bonuszahlungen nicht als Leistungsanreiz, sondern als bloße Missbrauchsbremse gewährt worden war.[311] Anders formuliert kompensierte die Hälfte der Vorstandsvergütung allein die größere Verhandlungsmacht, die der Vorstand dank der bis 1884 geltenden Regelungen genoss. Für die Frage der Anreizwirkung der Vorstandsvergütung verdeutlicht dieses Ergebnis, dass eine höhere Vergütung nicht unbedingt die Leistung steigert, sondern unter Umständen nur ein vorhandenes rechtliches Machtgefälle ausgleicht. Aus dieser Perspektive kritisieren Bayer und Burhop die neuere Tendenz zur erfolgsabhängigen Vorstandsvergütung als mögliches Anzeichen für den Niedergang des Anlegerschutzes.[312]

Wie jedes natürliche Experiment, beruht die Studie von Bayer und Burhop auf zwei Annahmen: Erstens, dass der betrachtete Strukturbruch – hier die Verschärfung der Organhaftung durch das Reformgesetz von 1884 – abrupt von außen ins Geschehen getreten ist, also exogen *wie* eine unabhängige Variable (dazu oben § 2 bei Fn. 67). Zweitens, dass der betrachtete Strukturbruch sich nur auf die relevante Zielgröße – hier die Höhe der Bonuszahlungen – ausgewirkt hat, nicht dagegen auf andere Einflussgrößen, die den vermuteten Ursachenzusammenhang verunreinigen. Bayer und Burhop begründen die erste ihrer beiden Annahmen damit, dass die Verschärfung der Organhaftung nur eine Nebenwirkung der eigentlich beabsichtigten Reform des Gründungsrechts dargestellt habe und nicht auf die Vergütungshöhe der Vorstände Einfluss nehmen sollte.[313] Zur zweiten An-

---

[310] *Bayer/Burhop*, Expl Econ Hist 2009, 464, 465: "Before 1884 each manager of a manufacturing firm (bank) received a profit share of about 3.8–4.6% (3.2% for banks), while after the reform, the profit share of a manager decreased to 0.9–1.9% in the manufacturing sector and 1.8% in the banking sector."; ähnl. *Bayer/Burhop*, sbr 2008, 378, 380.

[311] *Bayer/Burhop*, Expl Econ Hist 2009, 464, 479: "This incentive component was substituted for by better institutions under the new system of corporate governance introduced in Germany in 1884."

[312] *Bayer/Burhop*, Expl Econ Hist 2009, 464, 479; *Bayer/Burhop*, sbr 2008, 378, 397: "an increase in the pay-performance sensitivity […] might well reflect the worsening position of institutions of corporate governance."; auch *Stenzel*, Vorstandsvergütung 2012, 223, 231 folgert aus der empirischen Literatur, „dass Kontrollineffizienzen in der unternehmensinternen Corporate Governance-Struktur vergütungssteigernd wirken."

[313] *Bayer/Burhop*, Expl Econ Hist 2009, 464, 469 m.w.N.: "the reform in 1884 was intended to solve problems of incorporation […] not motivated by excessive management compensation. In fact, the legislators did not consider existing compensation schemes or problems

nahme stellen sie fest, dass die Reform von 1884 nur den gesellschaftsrechtlichen Vertragsrahmen änderte, nicht dagegen die Fähigkeit der betroffenen Vorstände oder die Wirtschaftslage allgemein.[314]

Diese Annahmen lassen sich freilich hinterfragen. Selbst wenn die schärfere Vorstandsregulierung nicht den vorrangigen Zweck der Reform von 1884 bildete, so war sie zumindest für informierte Marktteilnehmer (und spätestens ab der ersten Zirkulation des Gesetzentwurfs für jedermann) vorhersehbar. Das Gesetz hatte eine lange Vorgeschichte, von der preußischen Resolution 1876 und dem ersten Gesetzentwurf 1880 bis hin zum endgültigen Entwurf 1882 und seiner Einbringung in den Bundesrat 1883.[315] Demnach könnte es durchaus Unternehmen gegeben haben, die ihre Vergütungspraxis in Erwartung der kommenden Reform schon früher änderten. Dann aber hätte die Reform keinen abrupten Strukturbruch dargestellt, oder zumindest nicht im Jahr 1884. Dass die Studienautoren *trotzdem* einen Unterschied zwischen den Jahren vor und nach 1884 fanden, lässt zwei Deutungen zu – entweder dass der Gesamteffekt, den die Reform auf die Vergütungspraxis hatte, noch viel stärker gewesen sein muss, oder dass der festgestellte Unterschied vorrangig andere Umbrüche im selben Zeitraum widerspiegelte, wie zum Beispiel die gleichzeitig betriebene Sozialgesetzgebung Bismarcks. Obwohl sie also keinen ganz eindeutigen Ursachenrückschluss ermöglichen,[316] sind natürliche Experimente wie jenes von Bayer und Burhop doch eine der einfallsreichsten Varianten der empirischen Ursachenforschung, und erlauben bei maximaler äußerer Gültigkeit zumindest sehr überzeugende Ursachenschlüsse. Dafür haben sie einen gravierenden Nachteil: Der Forscher muss eine Situation finden, in der „von außen" die interessante Zielgröße (und nur sie) nennenswert beeinflusst worden ist. Das passiert nicht oft, mitunter nur einmal in hundert Jahren – nicht umsonst verwenden Bayer und Burhop Daten aus dem späten 19. Jahrhundert.[317] Solche Situationen abzuwarten oder zu suchen, ist meist aussichtslos, daher sind natürliche Experimente fast immer Gelegen-

---

with those in the run-up debate for the 1884 reform at any point".

[314] *Bayer/Burhop*, Expl Econ Hist 2009, 464, 466 ("legal reforms of corporate governance will typically not affect the abilities or other fixed characteristics of a manager") und 479 ("This reform enhanced corporate governance institutions, but it did not change the characteristics of managers or the influence of the business cycle").

[315] Zur Chronologie *Schubert* in: Schubert/Hommelhoff, Modernes AktR 1985, 1, 51 f.; zum gleich gelagerten Problem, das „Ereignis" für die Zwecke einer Ereignisstudie zu definieren, oben § 1 ab Fn. 308.

[316] So auch *Klick*, JITE 2010, 166: "the kinds of shocks that are large enough to change legal and political institutions are likely to change many different kinds of unobservable social phenomena, leading to the strong possibility of omitted variables biases".

[317] *Bayer/Burhop*, sbr 2008, 378, 380: "major reforms that qualify as such an experiment are naturally rare."; ebenso *Bayer/Burhop*, Expl Econ Hist 2009, 464, 465; vgl. oben Fn. 305.

heitsstudien,[318] die eine zufällig entstandene Datenstruktur opportunistisch nutzen.

Bayer und Burhop nutzten einen *zeitlichen* Bruch im Datensatz als Deutungsstrategie (*identification strategy*), genau wie die Ereignisstudie (oben § 1 G.II.). Ebenso lässt sich aber auch die unterschiedliche Entwicklung *räumlich* getrennter Gruppen vergleichen, die sich nur durch den Eintritt oder das Ausbleiben des jeweils relevanten Ereignisses unterscheiden (*difference-in-differences*).[319] Diese Deutungsstrategie wurde zu Fragen der Vorstandsvergütung oft eingesetzt,[320] birgt aber nicht nur zahlreiche methodische Fehlerquellen,[321] sondern hängt wiederum von der Überzeugungskraft der Annahme ab, dass die gewählten Vergleichsgruppen sich wirklich nur hinsichtlich des betrachteten Ereignisses unterscheiden. Diesem Einwand versucht eine weitere Deutungsstrategie zu begegnen, die gezielt solche Einzelfälle sucht und paarweise vergleicht (*matching*),[322] die sich im Vorliegen der vermuteten Ursache unterscheiden, aber ansonsten in möglichst allen Dimensionen identisch sind. Je besser das gelingen soll, desto aufwändiger wird diese Methode.

Letztlich steht und fällt jedes Quasi-Experiment mit der Überzeugungskraft seiner Deutungsstrategie.[323] Im schlechtesten Fall ist gegenüber dem Vorgehen einer Korrelationsstudie nicht viel gewonnen, im besten Fall ergeben sich Erkenntnisse von nahezu (labor)experimenteller Eindeutigkeit und gleichzeitig sehr hoher Übertragbarkeit.

e) Feldexperimente (Beispiel: Bandiera, Barankay & Rasul 2007)

Einen letzten Schritt von der Korrelationsstudie in Richtung des Laborexperiments macht das Feldexperiment, indem es einen realen Lebenssachverhalt gezielt in einer Weise manipuliert, die hohe innere Gültigkeit begünstigt. Dabei lassen sich drei Arten von Feldexperimenten unterschei-

---

[318] *Harrison/List*, J Econ Lit 2004, 1009, 1041: "The common feature of these experiments is serendipity".

[319] Beispiel bei *Braun/Eidenmüller u.a.*, ZHR 2013, 131, 139, sowie unten bei Fn. 344; technisch erläutert bei *Angrist/Pischke*, Econometrics 2009, 227 ff. (Kap. 5.2).

[320] Beispielstudien zitiert bei *Bayer/Burhop*, Expl Econ Hist 2009, 464, 466 Fn. 7.

[321] Vgl. etwa *Bertrand/Duflo/Mullainathan*, Q J Econ 2004, 249.

[322] *Epstein/Martin* in: Cane/Kritzer, Hdb Empirical 2010, 901, 916; *Diekmann*, Sozialforschung 2012, 359; *Legewie*, KZfSS 2012, 123, 132 f. und 148 f.; ausf. *Gangl/DiPrete*, KZfSS Sonderheft 2004, 396, 406 ff.; *Gensler/Skiera/Böhm*, JfB 2005, 37; *Ho/Rubin*, Ann Rev L Soc Sci 2011, 17, 28 f.; recht technisch *Angrist/Pischke*, Econometrics 2009, 69 ff. (Kap. 3.3.1).

[323] Vgl. auch *Harrison/List*, J Econ Lit 2004, 1009, 1014 ff. mit eingehenderen, wenn auch technischen, Erläuterungen.

den:[324] künstliche (*artifactual*) Feldexperimente, die Teilnehmer aus dem relevanten Zielkontext zur Teilnahme an einem regulären Laborexperiment einladen, eingebettete (*framed*) Feldexperimente, die direkt im Zielkontext durchgeführt werden, und natürliche (*natural*) Feldexperimente, deren Teilnehmer im Zielkontext ohne ihr Wissen einer von mehreren Versuchsbedingungen zugewiesen werden. Diese dritte Art von Feldexperimenten vermeidet ohne Weiteres den aus Feldstudien bekannten Hawthorne-Effekt (oben bei Fn. 241), wirft dafür jedoch forschungsethische Fragen auf (oben § 2 bei Fn. 95). Gleichwohl war sie Methode der Wahl für drei Ökonomen, die unlängst untersuchten, wie sich Vergütungsanreize für die Führungsriege auf die Produktivität ihres Unternehmens auswirken.[325]

Die Studie wurde auf der Plantage eines führenden britischen Obstproduzenten durchgeführt, deren Führungsriege – bestehend aus einem Geschäftsführer (COO) und zehn untergeordneten Personalverantwortlichen – die Ernte mit Hilfe angestellter Beerenpflücker zu verantworten hatte.[326] Dabei konnten die elf Unternehmensleiter die Produktivität ihrer 197 Akkordarbeiter auf zwei Wegen beeinflussen:[327] Zum einen konnte der Geschäftsführer Arbeiter einstellen, versetzen oder entlassen, und sich dabei fest auf das Überangebot von Saisonarbeitern vor Ort verlassen.[328] Zum anderen konnten die Personalverantwortlichen einzelne Arbeiter stärker oder weniger stark bei der Arbeit unterstützen.[329] Mithilfe von täglichen Aufzeichnungen über die Arbeitsleistung einzelner Arbeiter untersuchten die Studienautoren nunmehr die Auswirkungen eines finanziellen Mehranreizes für die Führungsriege auf Auswahl und Beaufsichtigung der Erntearbeiter. Dazu unterteilten sie die Haupterntesaison (vom 1. Mai bis zum 31. August) 2003 in zwei Zeitabschnitte von jeweils zwei Monaten Länge.[330] Im ersten Abschnitt erhielten die elf Unternehmensleiter ihr übliches Fest-

---

[324] *Harrison/List*, J Econ Lit 2004, 1009, 1014: “We recognize that any such taxonomy leaves gaps, and that certain studies may not fall neatly into our classification scheme.”; *List*, Sci 2008, 207; Bsp. auf www.fieldexperiments.com.

[325] *Bandiera/Barankay/Rasul*, Q J Econ 2007, 729, vgl. 741: “The COO and managers did not know that they were taking part in an experiment and that the data would be used for scientific research.”

[326] *Bandiera/Barankay/Rasul*, Q J Econ 2007, 729, 730 f., 733 f.

[327] *Bandiera/Barankay/Rasul*, Q J Econ 2007, 729, 730 (“as in most firms”), 743.

[328] *Bandiera/Barankay/Rasul*, Q J Econ 2007, 729, 731 f. bzw. 735: “individuals are hired seasonally from Eastern Europe, and they live on the farm for the duration of their stay. Importantly, there is typically an excess supply of bottom-tier workers, and work is offered on a causal [casual?] basis with no daily guarantee of employment.”

[329] *Bandiera/Barankay/Rasul*, Q J Econ 2007, 729, 735: “For example, by assigning her [i.e., the worker] to more plentiful rows [of fruit to pick] and removing her full crates quickly […] The effect of managerial effort on worker productivity can be substantial.”

[330] *Bandiera/Barankay/Rasul*, Q J Econ 2007, 729, 731, 740 f., 743.

gehalt, im zweiten Abschnitt *zusätzlich* einen täglichen Leistungsbonus, der von der durchschnittlichen Produktivität der jeweils betreuten Arbeiterschaft abhing.[331] Die Autoren stellten fest, dass die Produktivität der Arbeiter in der zweiten Hälfte der Erntesaison um durchschnittlich 21 % höher ausfiel, aber auch um 38 % stärker schwankte.[332] Das lag daran – wie die weitere Auswertung zeigte – dass das Führungspersonal einerseits vermehrt fähigere Arbeiter zur Arbeit anstellte,[333] sich andererseits aber auf die Unterstützung dieser fähigen Arbeiter so sehr konzentrierte, dass die übrigen Arbeiter vernachlässigt wurden und bisweilen sogar an Produktivität verloren.[334] Daraus schlossen die Autoren, dass zusätzliche Leistungsanreize für die Führungsriege zu einem unternehmensschädlichen Auseinanderdriften der Arbeiterschaft führen können.[335]

Als *Feld*experiment begünstigt die Studie von Bandiera und Kollegen zum einen die äußere Gültigkeit, da ihre Erkenntnisse auch auf andere Zeitfenster, Personen und womöglich sogar Erfolgsmaße (also U, T und O, oben bei Fn. 22) im untersuchten Kontext übertragen werden können – so sehr sogar, dass die Unternehmensführung des betroffenen Konzerns das von den Forschern untersuchte Bonusprogramm zum Erntejahr 2005 erneut einführte, um die Unternehmensgewinne zu steigern.[336] Die Übertragbarkeit auf andere *Kontexte* (S) hingegen wird dadurch erschwert, dass die vorliegende Studie (wie die meisten Feldexperimente) vollständig innerhalb *eines* Unternehmens stattfand. Zudem hatte dieses Unternehmen im Fall von Bandiera und Kollegen auch noch sehr eigenwillige Merkmale, beruht es doch auf saisonal beschränkter Akkordlohnbeschäftigung vorrangig für osteuropäische Studenten.[337] Die Autoren der Studie selbst erkannten die beschränkte Übertragbarkeit ihrer Erkenntnisse als Problem, da der

[331] *Bandiera/Barankay/Rasul*, Q J Econ 2007, 729, 731, näher 742: "Conditional on obtaining the bonus, managerial hourly earnings increased by 25 percent."

[332] *Bandiera/Barankay/Rasul*, Q J Econ 2007, 729, 732, 753 f.

[333] *Bandiera/Barankay/Rasul*, Q J Econ 2007, 729, 745: "when managers are paid performance bonuses, only 130 out of the 197 workers continue to pick. The remaining 67 workers are 'fired' from picking and either allocated to nonpicking tasks or left unemployed for some days."; ebenso auf 761.

[334] *Bandiera/Barankay/Rasul*, Q J Econ 2007, 729, 732: "the most able workers experience a significant increase in productivity while the productivity of other workers is not affected or even decreases. This suggests that […] after the introduction of performance pay, managers target their effort towards more able workers."

[335] *Bandiera/Barankay/Rasul*, Q J Econ 2007, 729, 732: "The introduction of managerial performance pay thus exacerbates earnings inequality" und weiter 771: "exacerbating inequality due to natural ability differences may be detrimental to the firms' long run performance. […] Whether firms trade off the benefits of incentive pay with these types of longrun effects when designing compensation schemes deserves further research."

[336] *Bandiera/Barankay/Rasul*, Q J Econ 2007, 729, 745 Fn. 12.

[337] *Bandiera/Barankay/Rasul*, Q J Econ 2007, 729, 735 Fn. 2.

Arbeitsmarkt für die Beerenernte ungewöhnlich abgeschlossen ist und die Pflücker ungewöhnlich isoliert voneinander arbeiten.[338]

Als Feld*experiment* begünstigt die Studie von Bandiera und Kollegen aber auch die innere Gültigkeit ihrer Erkenntnisse, da sie die relevanten Einflussgrößen in einem Maß kontrollieren konnte, das nur noch Laborexperimente übertreffen.[339] Daher gelten „gut gemachte Feldexperimente“ als das „beste, was an rechtlich relevanter Forschung überhaupt möglich ist.“[340] Dementsprechend konnte die vorliegende Studie – als einzige der hier vorgestellten – überhaupt einen Mechanismus benennen, durch den sich die höhere Vorstandsvergütung in höherer Unternehmensleistung niederschlägt. Statt die „Leistung“ des Vorstands nur als persönliche Leistung (wie in Laborexperimenten) oder nur als Unternehmensleistung (wie in Korrelationsstudien) zu operationalisieren, konnte das Feldexperiment den Weg nachverfolgen, auf dem sich Leistungsanreize für die Führungsriege durch das Unternehmen fortpflanzen.[341]

Andererseits illustriert die Studie von Bandiera und Kollegen auch die methodischen Herausforderungen, denen ein „gut gemachtes“ Feldexperiment begegnen muss. Beispielsweise, dass „die experimentelle Manipulation [.] schwer, oft sehr aufwendig und selten ganz perfekt gemäß einem Versuchsplan zu erreichen“ ist:[342] Aufgrund der Vertrautheit der Plantagenleiter miteinander konnten Bandiera und Kollegen ihnen nicht zur gleichen Zeit unterschiedliche Vergütungsanreize zu bieten,[343] deshalb verlegten die Autoren sich darauf, die verschiedenen Vergütungsanreize in verschiedenen Zeitabschnitten zu setzen. Dadurch aber stand zu befürchten, dass andere Entwicklungen innerhalb der Erntesaison den beobachteten Effekt verunreinigen könnten, deshalb verglichen die Autoren ihre Daten mit de-

---

338 *Bandiera/Barankay/Rasul*, Q J Econ 2007, 729, 769 f.: “inevitably entails some loss of generality because the firm we study, as any other, has unique features that influence the effect of managerial incentives on productivity.”

339 Dazu *Bandiera/Barankay/Rasul*, Q J Econ 2007, 729, 731: “we had full control over the timing of the change, the structure of managerial compensation, and the information provided to managers. Second, we observe the same workers and managers under both managerial incentive schemes and therefore control for time invariant sources of heterogeneity across workers, such as their ability, and across managers, such as their management style.”

340 *Lempert*, J Emp L Stud 2010, 907, 918; *Lawless/Robbennolt/Ulen*, Methods 2010, 31 f.: das “*controlled randomized field experiment*” sei der “gold standard”, aber “relatively infrequent, though not nonexistent, in the legal context.”

341 Zu Recht daher *Bandiera/Barankay/Rasul*, Q J Econ 2007, 729, 733: “we open the black box of behavior within the firm”.

342 So schon *Häcker* in: Bender, Tatsachenforschung 1972, 143, 155 zum Feldexperiment.

343 *Bandiera/Barankay/Rasul*, Q J Econ 2007, 729, 742: “each manager is aware of the compensation scheme offered to other managers. This raises the possibility of contamination effects if different managers were contemporaneously paid according to different compensation schemes.”

nen des darauffolgenden Jahres (2004), in dem keine Bonusregelung vorgesehen war, und untersuchten nur den Unterschied in den unterschiedlichen Entwicklungen (*difference-in-differences*, dazu schon oben bei Fn. 319).[344] Auf diese Weise gelang ihnen ein überzeugend eindeutiger Rückschluss auf die Ursachenzusammenhänge. Dennoch bleibt die Eindeutigkeit von Erkenntnissen aus Feldexperimenten immer in gewissem Umfang beschränkt: „In der natürlichen Umgebung lassen sich die Störvariablen ersichtlich nicht in vergleichbarer Weise kontrollieren."[345]

Schließlich illustriert die Studie von Bandiera und Kollegen ein weiteres Augenmerk bei der Rezeption empirischer Forschung, das bisher nur kurz (§ 1 bei Fn. 157) aufgetaucht ist: das Problem der Auftragsforschung. Die Studienautoren erwähnen nicht, wie und auf wessen Initiative sie den Zugang zur Obstplantage des nicht namentlich genannten Konzerns erhielten und unter welchen Bedingungen ihre Forschung stattfand. Immerhin aber hatten ihre Erkenntnisse substantiellen Wert für die Unternehmensrendite (vgl. oben bei Fn. 336) und wurde das Manuskript vom Konzern vor der Veröffentlichung überprüft,[346] daher ist die Frage sicher legitim, ob die Studienautoren ein Eigeninteresse an einem bestimmten Ausgang der Studie hatten. Soweit dafür keine konkreten Anhaltspunkte bestehen, ist ein Generalverdacht natürlich überzogen – schließlich muss das methodische Vorgehen überzeugen, nicht die wirtschaftlichen Hintergründe. Dennoch muss der Rezipient empirischer Forschung die hinter einer Studie stehenden Interessen reflektieren, soweit er sie erkennt.

3. Sind die Ergebnisse konsistent? Zur Meta-Metastudie von Mitchell 2012

Die bis hier dargestellte Vielfalt der Methoden wirft die Frage auf, ob sich damit überhaupt konsistente Erkenntnisse gewinnen lassen (sog. Triangulation[347]). Diese Frage beantworten Forschungssynthesen, die die an unterschiedlichen Enden des Spektrums gewonnenen Erkenntnisse vergleichen.

Da solche Synthesen meist die weit verbreiteten Zweifel an der laborexperimentellen Methode aufgreifen, untersuchen sie die Konsistenzfrage genau aus dieser Blickrichtung: Sie fragen danach, ob Laborexperimente „Wahrheit oder Wahn" produzieren, und stellen das Experiment allen anderen Methoden gegenüber.[348] Dabei ist das Problem bei genauerer Betrach-

---

[344] *Bandiera/Barankay/Rasul*, Q J Econ 2007, 729, 744 f., 756; allg. *Legewie*, KZfSS 2012, 123, 135 f. und 150.

[345] *Hussy/Schreier/Echterhoff*, Forschungsmethoden 2010, 135.

[346] *Bandiera/Barankay/Rasul*, Q J Econ 2007, 729: "This paper has been screened to ensure no confidential information is revealed."

[347] Anschaulich *Zeisel/Kaye*, Figures 1997, 69 ff. (Kap. 5).

[348] *Anderson/Lindsay/Bushman*, Curr Dir Psy Sci 1999, 3 mit dem hier frei übersetzten Untertitel "Truth or Triviality?"

tung symmetrisch: Erkenntnisse aus dem Feld weisen insgesamt keine höhere Gültigkeit auf als Laborexperimente – sie gewichten nur anders –, daher kann die Frage nur lauten, ob die Methoden übereinstimmende Erkenntnisse erzielen. Mehr kann auch die beste Forschungssynthese nicht feststellen, denn jeder Mangel an Übereinstimmung kann immer ebenso gut aus der mangelnden Übertragbarkeit von Labor-Erkenntnissen resultieren wie aus der mangelnden Eindeutigkeit von Feld-Erkenntnissen.

Für die Frage, ob Labor- und Feldstudien konsistente Ergebnisse erzielen, ist zunächst eine Meta-Metastudie aus dem Jahr 1999 relevant, die 21 bis dahin erschienene Metastudien mit insgesamt 38 Vergleichen zwischen Labor und Feld auswertete.[349] Eine jüngst veröffentlichte Erweiterung dieser Studie brachte es sogar auf 217 Labor-Feld-Vergleiche aus 82 Metastudien.[350] Diese derzeit umfangreichste Forschungsauswertung legt jeweils den Vergleich des Laborexperiments mit der am weitesten „entfernten“ Studienform zugrunde.[351] Dabei ergab sich, dass 86 % der verglichenen Ergebnisse *qualitativ* konsistent waren.[352] Bei genauerer Aufschlüsselung variierten die Ergebnisse stark zwischen verschiedenen psychologischen Teildisziplinen: In der Organisationspsychologie war der Grad der Übereinstimmung sehr hoch, in der Sozialpsychologie eher gering,[353] obwohl das Gefälle für manche Themen sogar genau umgekehrt verlief.[354] Selbst *innerhalb* der Sozialpsychologie jedenfalls stimmten die Erkenntnisse aus Labor und Feld noch in 73,7 % der Fälle qualitativ überein.[355] Als wichtigsten Einflussfaktor für die unterschiedlichen Übereinstimmungsraten identifiziert die Meta-Metastudie die Größe der jeweils dokumentierten Effekte, was den Wert der vierten Rezeptionsregel (oben § 2 B.IV.) belegt.[356]

---

[349] *Anderson/Lindsay/Bushman*, Curr Dir Psy Sci 1999, 3, 5 f.

[350] *Mitchell*, Persp Psy Sci 2012, 109; zum Autor *Mitchell*, Virg J 2008, 28.

[351] *Mitchell*, Persp Psy Sci 2012, 109, 110: “comparison of laboratory effects to […] conditions that differ most from the laboratory conditions because these research settings possess the least ‘proximal similarity’ to the laboratory and thus are likely to raise the greatest generalizability concerns.”

[352] *Mitchell*, Persp Psy Sci 2012, 109, 111: “overall, 30 of 215 laboratory effects changed signs (14%).”

[353] *Mitchell*, Persp Psy Sci 2012, 109, 111: “Laboratory and field effects from I-O psychology correlate very highly (r = .89, n = 72, 95% CI [.83, .93]), whereas laboratory and field effects from social psychology show a lower correlation (r = .53, n = 80, 95% CI [.35, .67]).”; gerade die Organisationspsychologie allerdings war nach dem Forschungsbericht von *Dipboye/Flanagan*, Am Psy’st 1979, 141 sehr beschränkt in ihrer Feldforschung.

[354] Bspw. *Mitchell*, Persp Psy Sci 2012, 109, 113: “leadership studies within I-O psychology were less predictive than leadership studies within social psychology (r = .63 […] vs. r = .93 […]).”

[355] *Mitchell*, Persp Psy Sci 2012, 109, 112: “Twenty-one of 80 (26.3%) laboratory effects from social psychology changed signs between research settings”.

Zusammenfassend lässt sich also feststellen, dass die Labor- und Feldmethoden in der Mehrzahl der Fälle qualitativ konsistente Ergebnisse erbringen, dass aber je nach Forschungsfrage und Effektgröße durchaus Zweifel an der (äußeren oder inneren) Gültigkeit vieler (Labor- oder Feld-) Studien angebracht sind. Auch aus diesem Grund empfiehlt sich der Rückgriff auf Forschungssynthesen wo immer möglich (vgl. oben § 2 B.VII.) und der kritische Umgang mit einzelnen Primärstudien, die noch nicht mit anderen Studienarten und -methoden repliziert wurden.[357]

V. Ein Fazit zur Vorstandsvergütung?

Wie eingangs erwähnt, ging es hier weniger um die Frage der Vorstandsvergütung als vielmehr um die Illustration verschiedener Methoden, mit denen man diese Frage empirisch angehen kann. Die Höhe der Erfolgsvergütung ist ja nur einer von vielen Aspekten, und die hier vorgestellten Studien sind auch nur einige von vielen, die sich mit diesem und anderen Aspekten befassen.[358] Dennoch deuten sie einige Schlussfolgerungen an, die abschließend kurz zusammengefasst werden sollen:

Die Korrelationsstudie von Kaplan und Rauh findet einen statistischen Zusammenhang zwischen der Vergütungshöhe eines US-amerikanischen Vorstands und der Kapitalrendite seines Unternehmens: Während das bestbezahlte Fünftel der Vorstände eine im Vergleich zum Branchendurchschnitt 61 % höhere Kapitalrendite erzielt, erwirtschaftet das schlechtestbezahlte Fünftel der Vorstände Minderrenditen von 19 %. Ob die Leistung mit der Vergütung steigt oder die Vergütung mit der Leistung, muss die Studie aufgrund ihrer beobachtenden Ausrichtung offenlassen.

Die Feldstudie von Gardner, Van Dyne und Pierce findet – über Vorstände und Angestellte hinweg – einen schwachen statistischen Zusammenhang zwischen dem Vorjahresgehalt eines Mitarbeiters und seiner diesjährigen Personalbewertung: Sie bedingen einander zu 4 %. Ob die Personalbewertungen aber die tatsächliche Arbeitsleistung widerspiegeln oder unmittelbar durch die Vorjahresvergütung beeinflusst werden, oder ob der statisti-

---

[356] So auch *Mitchell*, Persp Psy Sci 2012, 109, 113 f., vgl. aber 115: "Determining the mix of factors responsible for the observed variations in external validity will require further research."

[357] *Häcker* in: Bender, Tatsachenforschung 1972, 143, 155: „so ist zu empfehlen, sowohl Feld- als auch Laboratoriumsexperimente durchzuführen und die Ergebnisse in einer Art Gegenvalidierung zu kontrollieren."; *Stanovich*, Psychology 1998, 109: "replication goes a long way toward answering the What-if-it-only-holds-in-Boston? complaint." und weiter 120 ff. zum "principle of converging evidence".

[358] Vgl. z.B. die "review of reviews" von *Durant/Kramer u.a.*, Publ Admin Rev 2006, 505 m. w. Nachw. und Anknüpfungspunkten, die hier nicht aufgegriffen werden konnten.

sche Zusammenhang nicht gänzlich auf dritte Faktoren zurückgeht, kann die Studie nicht beantworten.

Die Befragungsstudie von Pepper, Gore und Crossman findet unter britischen Führungskräften Ansichten, aus denen sich ergibt, dass die Vorstandsvergütung nur in einem Bereich mittlerer Höhe leistungssteigernd wirkt, dass eine zu geringe oder zu hohe Vergütung hingegen die Leistungsbereitschaft der Führungskraft zerstören kann. Ob die von den Führungskräften geäußerten Ansichten zutreffen oder auch nur in ehrlicher Absicht geäußert wurden, ist dagegen unklar.

Die Vignettenstudie von Weibel, Rost und Osterloh findet unter deutschen Berufstätigen mit Führungsverantwortung eine Steigerung der Leistungsmotivation durch variable Vergütung, die aber in einer Wechselwirkung mit der intrinsischen Motivation steht: Je höher die anfängliche Begeisterung und Einsatzbereitschaft, desto eher verringert erfolgsabhängige Vergütung die Leistungsmotivation, statt sie zu erhöhen. Ob aber die Leistungsmotivation, die die Teilnehmer bekundeten, ihrer tatsächlichen Leistungsbereitschaft in der Praxis entspricht, kann die Studie nicht klären.

Das Quasi-Experiment von Bayer und Burhop findet einen Unterschied zwischen der Höhe der Beteiligung deutscher Vorstände am Unternehmensgewinn vor und nach der Aktienrechtsreform von 1884, und schließt daraus, dass etwa die Hälfte der damaligen Vorstandsboni ausschließlich zum Ausgleich des Machtgefälles gegenüber den Aktionären diente und keine Anreizwirkung entfaltete. Inwieweit diese Erkenntnisse auf sonstige Vergütungsbestandteile und auf das heutige Aktienrecht übertragbar sind, kann die Studie allerdings nicht beantworten.

Das Feldexperiment von Bandiera, Barankay und Rasul findet eine verbesserte Auswahl und Unterstützung produktiver Mitarbeiter durch Unternehmensleiter, die einen höheren finanziellen Anreiz erhalten – zugleich aber eine Vernachlässigung der weniger produktiven Kräfte, die den Unternehmensfrieden stören und den Produktivitätsgewinn womöglich sogar aufzehren kann. Ob das in stärker hierarchisierten Unternehmen ohne saisonalen Produktionsbetrieb und mit größerer wechselseitiger Abhängigkeit der Arbeitnehmer ebenso passiert, kann die Studie nicht beantworten.

Das Laborexperiment von Ariely, Gneezy, Loewenstein und Mazar findet sowohl unter indischen Dorfbewohnern als auch unter US-amerikanischen Studenten einen relativen Leistungsabfall bei besonders hohen Anreizen zur Lösung kognitiv anspruchsvoller Aufgaben und schließt daraus, dass zu hohe Vergütungen durch übermäßigen Leistungsdruck zum Versagen führen können. Ob das auch auf Vergütung und Aufgaben des Vorstands übertragbar ist, lässt sich nicht sicher sagen.

In der Synopse jeder dieser sieben Studien habe ich ihr wesentliches Ergebnis dargestellt, zugleich aber auch ihre wesentliche Beschränkung angesprochen. Das soll verdeutlichen, dass keine empirische Herangehens-

weise pauschal besser oder schlechter geeignet ist, zur Beantwortung einer bestimmten Frage beizutragen.[359] Es gibt keine Erkenntnis von optimaler Gültigkeit, sondern nur einen Kompromiss aus unterschiedlichen Gültigkeiten. Allgemeine Erkenntnisse über die von vielen Studien zugrunde gelegte Fragestellung ergibt erst die anschließende Forschungssynthese.

Die für die vorliegende Frage derzeit wohl aktuellste[360] Forschungssynthese ist eine Metastudie über 75 Primärstudien zur Erfolgsvergütung (*pay-for-performance*).[361] Sie kommt zu dem Ergebnis, dass gerade einmal 0,64 % des Unternehmenserfolgs auf die Höhe des Vorstandsgehalts zurückzuführen sind;[362] der 1950 noch bestehende Zusammenhang von Erfolgsvergütung und Unternehmenserfolg sei mittlerweile bis zur Unkenntlichkeit geschrumpft.[363] Deshalb wirke sich eine höhere Gesamtvergütung sogar negativ auf den erwarteten Unternehmenserfolg aus[364] und sei insbesondere die variable Vergütung von Vorstandsmitgliedern inzwischen kontraproduktiv.[365] Zu einem ähnlichen Ergebnis kam schon ein zwei Jahre älterer Forschungsbericht unter Auswertung von 99 Primärstudien.[366]

Um die Übertragbarkeit dieser Erkenntnisse zu beurteilen, ist natürlich zu beachten, dass die meisten Primärstudien erstens aus den USA stammen, also deren kulturelle und institutionelle Besonderheiten reflektieren,[367] und zweitens mit öffentlichen Kapitalmarktdaten durchgeführt wurden, also die Besonderheiten börsennotierter Aktiengesellschaften reflek-

---

[359] Treffend *Henry*, Psy Inq 2008, 49, 66: “each methodology by itself is flawed in its own way and cannot definitively reveal the nature of any psychological phenomenon by itself. However, each methodology also carries with it certain truths, and when those truths converge to tell a coherent story, we can make more confident claims”.

[360] Zu älteren *Rost/Osterloh*, sbr 2009, 119, 132 Fn. 14; einen neuen Überblick über einschlägige Forschungssynthesen geben *Jacquart/Armstrong*, Interf 2013, 580, 582 ff.

[361] *Rost/Osterloh*, sbr 2009, 119, 130: die 75 Primärstudien umfassen 123.797 Unternehmen; deutschsprachige Urfassung verfügbar als *Rost/Osterloh*, Managementmode 2007.

[362] *Rost/Osterloh*, sbr 2009, 119, 132: “in other words a negligible amount. Previous investigations find similar results.” m.w.N. in Fn. 14.

[363] *Rost/Osterloh*, sbr 2009, 119, 134: “in 1950 a variable CEO income increased the firm performance at d = 0.21. […] Nowadays, salary and performance are only linked to each other at d = 0.05; an almost nonexistent link.”

[364] *Rost/Osterloh*, sbr 2009, 119, 137 f.: “We even find a negative correlation between total executive compensation and expected firm performance […] Thus, higher amounts of executive compensations do not necessarily increase shareholder wealth, but have the potential to decrease shareholder wealth.”

[365] *Rost/Osterloh*, sbr 2009, 119, 140: “Overall, the results of this study show that Pay-for-Performance has counterproductive effects”.

[366] *Devers/Cannella u.a.*, J Mgmt 2007, 1016, 1028: “In sum, the research evidence to date strongly supports the conclusion that executives use incentive compensation in ways that benefit themselves at the expense of shareholders.”, aber weiter 1032: “the unique elements of executive pay exhibit more complex influences than commonly understood […] simplistic conceptualizations of executive risk preferences and individual pay elements are incomplete.”

tieren. Zudem erlaubt die Metastudie keine Aussagen darüber, welche Vorstandsvergütung im *konkreten Einzelfall* leistungssteigernde oder leistungsmindernde Wirkung entfaltet. Jedenfalls aber zwingt sie den Aufsichtsrat, sich bei der Vereinbarung einer Vergütung nicht auf das Prinzip „viel hilft viel" zu verlassen und insbesondere auch variable Vergütungsbestandteile konkret zu begründen. „Es spricht viel dafür, dass Wissenschaft und Praxis Anreizsysteme [in Form der variablen Vergütung] überschätzen",[368] daher genügt der pauschale Verweis auf die vermeintliche Angleichung von Unternehmens- und Vorstandsinteresse nach den Erkenntnissen der empirischen Forschung nicht (mehr).[369]

---

[367] Vgl. auch *Hupka*, Vergütungsvotum 2012, 41 f., 333; in ähnlichem Zusammenhang *Fleischer*, ZGR 2008, 185, 189: „Nun wird man die US-amerikanischen Daten schon wegen der ganz unterschiedlichen Corporate-Governance-Struktur nicht maßstabsgetreu auf die hiesigen Verhältnisse übertragen dürfen, doch bleibt ein gewisses Unbehagen, wenn Gesetzesvorschläge […] auf gänzlich ungesicherter Tatsachenbasis unterbreitet werden."

[368] *Stenzel*, Vorstandsvergütung 2012, 227.

[369] Konstruktive Verbesserungsvorschläge macht *Stenzel*, Vorstandsvergütung 2012, 227 f.

§ 4

Hält das Kollegialprinzip, was es verspricht?

„Das Kollegialitätsprinzip versucht […] eine Art von Synthese der Fach*spezialisten* zu einer kollektiven Einheit. Mit welchem Erfolg, ist nicht allgemein auszumachen."[1]
– Max Weber 1922

„In welcher Weise sich der vermutete Leistungsvorteil der Gruppe tatsächlich verwirklicht, ist […] bislang nicht systematisch untersucht worden. Zweifel sind jedenfalls angebracht."[2]
– Jörg Berkemann 1988

Ein vom Gesetzgeber oft und mit nonchalanter Selbstverständlichkeit herangezogener Argumentationstopos ist „das Kollegialprinzip". So wurde in der Aktienrechtsreform 1965 das überkommene Alleinentscheidungsrecht des Vorstandsvorsitzenden mit der Begründung abgeschafft, es „verstößt gegen das Kollegialprinzip",[3] und auch in jüngerer Zeit beruft sich der Gesetzgeber sporadisch auf dieses Prinzip, etwa bezogen auf den Vorstand einer Anstalt des öffentlichen Rechts,[4] oder auf die Senate am Oberlandesgericht.[5] Was beinhaltet dieses Prinzip? Warum soll es beispielsweise für den Vorstand einer Aktiengesellschaft gelten? Das hat der Gesetzgeber zu keiner Zeit erläutert; offenbar erschien es ihm allzu offensichtlich. Im vorliegenden Kapitel widme ich mich den aufgeworfenen Fragen und lege dabei besonderes Augenmerk auf den Beitrag, den die empirische Forschung zu ihrer Beantwortung leisten kann.

Dabei folge ich grob den vier Schritten, die im Einleitungskapitel für die Rezeption empirischer Forschung entwickelt wurden (§ 1 E.II.), indem

[1] *Weber*, Wirtschaft 1922, 674.

[2] *Berkemann*, KritV 1988, 29, 51 (zur Rechtsfortbildung in Richterkollegien).

[3] Begr. § 77 AktG-RegE 1965, BT-Drs. 4/171 in: *Kropff*, AktG 1965, 99.

[4] Begr. § 15 II Entwurf eines Gesetzes zur Errichtung einer Deutschen Arzneimittel- und Medizinprodukteagentur, BT-Drs. 16/4374 v. 23.2.2007, 26; vgl. auch *Kluth/Goltz*, GewArch 2003, 265 zu den Handwerkskammern.

[5] Begr. Art. 1 Nr. 3 Entwurf eines Gesetzes zur Einführung erstinstanzlicher Zuständigkeiten des Oberlandesgerichts in aktienrechtlichen Streitigkeiten, BT-Drs. 16/9020 v. 30.4.2008, 16; Begr. Art. 2 § 568 und Gegenäußerung Nr. 6 zum Entwurf eines Gesetzes zur Reform des Zivilprozesses, BT-Drs. 14/4722 v. 24.11.2000, 111, 157.

ich zuerst die im Titel dieses Kapitels bezeichnete Frage – was der Topos „Kollegialprinzip“ dem Recht verspricht – dogmatisch ausformuliere und ihre empirischen Aspekte herausarbeite (A.), zweitens die empirischen Forschungsrichtungen sichte, die für diese Aspekte relevant sein könnten (B.), drittens die Erkenntnisse dieser Forschung darstelle und kritisch würdige (C.) und abschließend den Kreis zurück zur Rechtsfrage schließe und aus der Rezeption der empirischen Forschung Schlussfolgerungen zur Rechtsdogmatik des Kollegialprinzips und zu seinem rechtspolitischen Potential ziehe (D.).

Da die vorliegende Arbeit ihre Beispiele vor allem im Gesellschaftsrecht sucht, wird dieses Rechtsgebiet einerseits den Rahmen bilden, innerhalb dessen auch andere Erscheinungen des Kollegialprinzips in die Betrachtung einbezogen werden (unten § 4 A.I.2.), andererseits aber auch die Grenze der aus der empirischen Betrachtung gezogenen Schlussfolgerungen; ob und inwieweit die empirischen Erkenntnisse auch andere Rechtsgebiete befruchten können, sei den kompetenteren Fachtexten zu diesen Rechtsgebieten vorbehalten.

A. Dogmatische Annäherung an die Frage

Das Kollegialprinzip taucht weder im Gesetz[6] noch in den bisher herausgearbeiteten Prinzipien des Gesellschaftsrechts auf.[7] In der Tat stellt es wohl gar kein Wertungsprinzip nach herkömmlichem Verständnis dar, sondern eher einen funktionsbestimmten Rechtsbegriff.[8] Im Schrifttum jedenfalls „herrscht hinsichtlich des genauen Inhalts und der Herleitung des Kollegialprinzips eine gewisse Unklarheit.“[9]

Zur Klärung von Inhalt und Herleitung rechtlicher Begriffe verwendet die Rechtsdogmatik die Methode der Auslegung, insbesondere nach Wortlaut, Systematik, subjektiven und objektiven Zwecken. Die Auslegung im herkömmlichen Sinn beschränkt sich zwar auf die Gesetzesanwendung,[10] doch bei näherer Betrachtung „sind grundsätzlich *alle* Rechtstexte der Auslegung sowohl fähig wie bedürftig“.[11] Die Auslegung als allgemeine her-

---

[6] Die Volltextsuche auf www.gesetze-im-internet.de/volltextsuche.html ergab bis zum 22.2.2014 keine Treffer.

[7] Vgl. *Bydlinski*, Prinzipien 1996, 469 ff. m.Verw. auf *Wiedemann*, GftsR I 1980, 357 ff. (§§ 7–11).

[8] Dazu *Larenz/Canaris*, Methodenlehre 2008, 302 ff. (Prinzipien), 310 ff. (funktionsbestimmte Rechtsbegriffe).

[9] *Beckert*, Personalisierte Leitung 2009, 27.

[10] Vgl. nur *Larenz/Canaris*, Methodenlehre 2008, 91 ff., 133 ff. (Kap. 4).

[11] *Larenz/Canaris*, Methodenlehre 2008, 26.

meneutische Methode der Textinterpretation dient dem „Verstehen von sprachlichen Äußerungen“ überhaupt,[12] kann also auch für nicht normierte Rechtsbegriffe – wie „das Kollegialprinzip“ – fruchtbar gemacht werden. Dabei ist zwar eine etwas andere Schwerpunktsetzung geboten, weshalb der nachfolgende Aufbau nicht unmittelbar den kanonischen Auslegungsmethoden folgt und auch nur eine dogmatische „Konkretisierung“ anstelle einer „Auslegung“ des Kollegialprinzips anstrebt. Der zugrundeliegende Anspruch ist allerdings derselbe, daher orientiert sich die Darstellung maßgeblich an den Kategorien der grammatischen (I.1.), systematischen (I.2.) sowie rechtsvergleichenden (I.3.) Interpretation und fragt schließlich in rechtsvergleichender Perspektive nach dem Sinn und Zweck des Kollegialprinzips (II.), um die ihm zugrundeliegenden empirischen Annahmen herauszuarbeiten (IV.).

I. Konkretisierung und Bestandsaufnahme

Obwohl das „Kollegialprinzip“ gesetzlich nicht normiert ist, stellt § 2 I BGremBG immerhin „kollegiale Organe“[13] in eine Reihe mit Vorständen und Aufsichtsräten, indem er „Gremien“ im Sinne des Gesetzes definiert als: „Vorstände, Beiräte, Kommissionen, Ausschüsse, Verwaltungs- und Aufsichtsräte, kollegiale Organe und vergleichbare Gruppierungen“.[14] Für die Frage, ob der Gesetzgeber Vorstände und Aufsichtsräte als kollegiale Organe ansieht, liegen damit zwei Deutungen nahe: Einerseits der Umkehrschluss, dass es ihrer ausdrücklichen Erwähnung nicht bedurft hätte, wenn sie bereits vom Begriff der „kollegialen Organe“ erfasst wären, andererseits das Verständnis der Aufzählung als phänomenologische Liste von Regelbeispielen, die weder erschöpfend noch klar abgrenzbar gemeint waren. Für letzteres spricht, dass innerhalb der Vorschrift die „kollegialen Organe“ durch ein „und“ mit den „vergleichbaren Gruppierungen“ verknüpft sind, die ersichtlich eine Auffangkategorie am Ende der Aufzählung bilden. Da eine solche Auffangkategorie alle früher genannten Begriffe umfassen muss, wären Vorstand und Aufsichtsrat in dieser Lesart nur ein namentlich genanntes Beispiel für die „kollegialen Organe und vergleichbaren Gruppierungen“. Letztlich ist der Wortlaut insoweit unklar, und auch die Gesetzesbegründung trägt zur Klärung nichts bei, weil sie ihn nur wörtlich wiederholt.[15] Auch eine Annäherung über Sinn und Zweck führt bei § 2

[12] *Larenz/Canaris*, Methodenlehre 2008, 25.

[13] Im übrigen erwähnen nur noch § 191f II BRAO das „Kollegialorgan“ und § 88 VwVfG „kollegiale Einrichtungen“.

[14] § 2 I BGremBG, seinerseits Art. 11 Zweites Gleichberechtigungsgesetz (2. GleiBG) v. 24.6.1994, BGBl. I 1406.

[15] Begr. Art. 12 § 2 I GleiBG2-RegE, BT-Drs. 12/5468 v. 21.7.1993, 49.

BGremBG nicht weiter, da das gesamte Gesetz Gleichstellungszwecken dient und der materielle Gehalt der Definitionsvorschrift offenbar nicht weiter reflektiert wurde.

Letztlich legt die Einordnung „kollegialer Organe“ unter die „Gremien“ einen Blick in das Verwaltungsorganisationsrecht nahe (dazu unten 2. und II.2.). Da aber das BGremBG ansonsten nicht weiterführt, kann die Konkretisierung des „Kollegialprinzips“ nur bei diesem Begriff selbst beginnen.

1. Wortlaut: Organpluralität und Organkollektivität

Das Wort „Kollegialprinzip“ ist ein Kompositum zweier lateinischer Lehnwörter: „kollegial“ von lat. *collega*, Amtsgenosse, und „Prinzip“ von lat. *principium*, Grund, Ursprung, Anfang; beide Ausgangsbegriffe sind ihrerseits wieder Komposita: *collega* von *com* (mit) und *legare* (ernennen, entsenden); *principium* von *primus* (erste), und *capere* (fassen, ergreifen).[16]

Damit lässt sich die etymologische Wortbedeutung grob umreißen als „Grundsatz von der gemeinsamen Ernennung“; zwei Merkmale des Kollegialprinzips dürften demnach sein:

1. Mehrere Menschen
2. handeln gemeinsam…

Bleibt man an diesem Punkt stehen, erhält man eine erste Definition des „Kollegialprinzips“, die sich auf Verbandsorgane beziehen lässt: Organe werden mit mehreren Mitgliedern besetzt. Das „Kollegialprinzip“ wird dann also als *Grundsatz der Organpluralität* verstanden. Diese Deutung ist zwar sehr weit, aber durchaus gebräuchlich.[17] Weiter verbreitet ist allerdings ein engeres Begriffsverständnis, das auch berücksichtigt, wie die gemeinsam bestellten Organmitglieder zusammenwirken. So bezeichnet das „Kollegialprinzip“ allgemeinsprachlich die „gemeinsame Beschlussfassung von gleichberechtigten Personen“,[18] woraus zwei weitere Merkmale abgeleitet werden können:

---

[16] *Harper*, Etymology 2014, Stichworte “colleague”, “principle” und “prince”; *Köbler*, Etymologisches Wb. 1995, S. 222 („Kollege“ / „kollegial“), 314 („Prinzip“); anders o. Nachw. *Wettich*, Vorstandsorganisation 2008, 36 f.

[17] Bspw. *Kalss* in: Doralt/Nowotny/Kalss, öAktG 2012, § 86 Rn. 8: „Kollegialorgan: Da das Gesetz mindestens drei Personen für die Besetzung des Aufsichtsrats vorschreibt, handelt der Aufsichtsrat immer als Gremium“.

[18] Duden online, www.duden.de/zitieren/10042528/2.0, abgerufen am 23.6.2014; *Dudenredaktion*, Fremdwörterbuch 2000, 717 verweist auf das „Kollegialsystem“ als sinnverwandtes Konzept und definiert es als „gemeinsame Verwaltung u. Beschlussfassung [von gleichberechtigten Personen in einer Behörde]“ (Kommentar im Original).

3. … und gleichberechtigt

4. durch Beschlussfassung…

Die so gewonnene allgemeinsprachliche Begriffsbestimmung ist am rechtswissenschaftlichen Sprachgebrauch näher zu präzisieren. Das Kollegialprinzip taucht allerdings in ganz verschiedenen Rechtsgebieten auf (dazu gleich 2.), deren Sprachgebrauch nicht unbedingt übereinstimmt. Auch innerhalb eines Rechtsgebiets finden sich erhebliche Unterschiede, vor allem im hier relevanten Gesellschaftsrecht: Manche Gesellschaftsrechtler setzen das Kollegialprinzip mit dem Prinzip der Gesamtgeschäftsführung gleich,[19] andere mit dem Grundsatz der Gesamtvertretung[20] und wieder andere unterscheiden stattdessen die Gesamtgeschäftsführung als „strenge Variante des Kollegialprinzips" von den unabdingbaren Mindestzuständigkeiten des Leitungsorgans als „abgemildertem Kollegialprinzip".[21] Diese Begriffsverwirrung dürfte zum größten Teil auf sprachlicher Sorglosigkeit beruhen, denn bei Licht betrachtet scheint weitgehend Einigkeit zu herrschen,[22] dass das Kollegialprinzip zwei Aspekte umfasst: Gleichberechtigung und Gleichverpflichtung.[23] So sei einerseits „mit der Einrichtung eines Kollegialorgans stets die Vorstellung verbunden, dass es sich um ein Kollegium von jedenfalls im Grundsatz gleichberechtigten Mitgliedern handelt."[24] Dieser Aspekt des Kollegialprinzips dürfte hinter der Aussage des Gesetzgebers von 1965 stehen (oben bei Fn. 3), das Alleinentscheidungsrecht des Vorstandsvorsitzenden verstoße gegen das Kollegialprinzip. Andererseits beinhalte das Kollegialprinzip „auch eine Gleichverpflichtung. [… Es sind] alle Organwalter in gleichem Maße verantwortlich für das gesamte Organhandeln."[25] Damit ergibt sich ein fünftes Begriffsmerkmal:

5. … die alle Mitglieder gleichermaßen verpflichtet.

---

[19] *Bürkle*, AG 2012, 232; *Seibt* in: Lutter/Schmidt, AktG 2010, § 77 Rn. 2, 7; wohl auch *Dauner-Lieb* in: Henssler/Strohn, GftsR 2014, § 77 AktG Rn. 1, aber anders *dies.* in: Henssler/Strohn, GftsR 2014, § 78 AktG Rn. 1: Gesamtgeschäftsführung und Gesamtvertretung als Ausprägungen des Kollegialprinzips.

[20] *H.F. Müller* in: Fleischer/Goette, MK-GmbHG 2011, § 68 Rn. 3, 4.

[21] *Von Werder* in: Ringleb/Kremer u.a., DCGK 2014, Rn. 96.

[22] Andernorts prägnant *Graumann*, ZGR 2011, 293, 297: „Man verstand und versteht sich, obwohl nicht im Letzten geklärt ist, worüber man redet […] Das ist übrigens nichts Ungewöhnliches."

[23] *Hoffmann-Becking*, NZG 2003, 745, 746 f.; ähnl. *Fleischer*, NZG 2003, 449, 458; *Wicke*, NJW 2007, 3755.

[24] *Hoffmann-Becking*, NZG 2003, 745, 746 m. Verw. auf Verwaltungsorganisations-, Staats-, Vereins- und Kapitalgesellschaftsrecht; aus US-amerikanischer Sicht schon *Haft*, Mich L Rev 1981, 1, 19 f.

[25] *Hoffmann-Becking*, NZG 2003, 745, 746 f.; *Beckert*, Personalisierte Leitung 2009, 28.

Anstelle der Bezeichnung „Gleichverpflichtung“ findet man häufiger den Begriff der „Gesamtverantwortung“.[26] Diese Wortwahl ist allerdings wiederum missverständlich: Erstens ermöglicht sie keine genaue Abgrenzung der zwei Aspekte „Berechtigung“ und „Verpflichtung“, sondern vermischt beide fast zwangsläufig.[27] Zweitens wird „die Gesamtverantwortung“ ähnlich sorglos mit dritten Begriffen gleichgesetzt wie schon „das Kollegialprinzip“.[28] Daher betone ich hier die *verpflichtende* Dimension der Gesamtverantwortung (d.h. den Haftungsaspekt).

Damit ergibt sich als endgültige Definition: Kollegialprinzip ist die grundsätzlich gemeinsame und gleichberechtigte Beschlusstätigkeit mehrerer Entscheidungsträger, die dadurch jeweils gleiche Verpflichtungen eingehen.[29] Dieses Verständnis des Kollegialprinzips lässt sich kürzer als *Grundsatz der Organkollektivität* bezeichnen.

Sind damit zwei verschiedene Deutungen des „Kollegialprinzips“ aufgezeigt (Organpluralität und Organkollektivität), bleibt deren Verhältnis untereinander zu klären. Das scheint zunächst trivial: Man könnte versucht sein, ein Spezialitätsverhältnis dergestalt anzunehmen, dass Organkollektivität stets Organpluralität voraussetzt. Schließlich ist die Beschlussfassung mehrerer (Organkollektivität) nicht denkbar, ohne dass überhaupt mehrere bestellt sind (Organpluralität). Wie aber ist der Fall zu bewerten, dass rechtlich keine Organpluralität vorgeschrieben, wohl aber Organkollektivität angeordnet ist, *wenn und soweit* Organe trotzdem plural besetzt werden? Das ist genau der Fall der §§ 76 f. AktG, und er lässt sich mit keiner der bisherigen Definitionen des „Kollegialprinzips“ sprachlich unzweideutig erfassen. Stattdessen wurden Organkollegialität und Organkollektivität oft ineins gesetzt,[30] und kaum je danach gefragt, ob „das Kollegialprinzip“ eigentlich Wunsch oder Wirklichkeit bezeichnen soll.

Eine differenzierende Herangehensweise muss deshalb zum einen die Grundsätze der Organpluralität und Organkollektivität voneinander unterscheiden, zum anderen aber auch danach differenzieren, ob der jeweilige „Grundsatz“ normativ oder deskriptiv formuliert ist, also rechtliche Anfor-

---

[26] Etwa bei *Rottnauer*, NZG 2000, 414, 416; *Fleischer*, NZG 2003, 449, 458 und *Wicke*, NJW 2007, 3755.

[27] Sehr deutlich bei *Fleischer*, NZG 2003, 449, 450 und 458: das Kollegialprinzip sei „Gesamtverantwortung und Gleichberechtigung“, zugleich aber sei die gleichberechtigte Gesamtleitung Teil der Gesamtverantwortung.

[28] Vgl. etwa *Hanau*, ZGR 1983, 346, 369; *Haas/Ziemons* in: Michalski, GmbHG 2010, § 43 Rn. 153 (jeweils: Gesamtverantwortung = Gesamtgeschäftsführung).

[29] Ähnl. *Weber*, Wirtschaft 1922, 159, der in Abgrenzung zur „Kassationskollegialität“ für die „Leistungskollegialität“ verlangte, dass „nach vorgängiger Beratung und Abstimmung […] nicht ein einzelner, sondern eine Mehrheit von einzelnen zusammenwirken muss, damit eine bindende Verfügung zustande kommt.“

[30] Etwa bei *Dose*, ZGR 1973, 300, 308.

derung oder faktische Handhabung beschreibt.[31] Nach dieser Differenzierung sind rein rechnerisch $(2^2)^2 = 16$ Fälle denkbar, deren fünf anschaulichste wie folgt lauten:

Der *Einzelkaufmann* (§ 1 HGB) ist weder rechtlich noch faktisch plural und weder rechtlich noch faktisch kollektiv.

Der übliche *Vorstand* von Verein oder Aktiengesellschaft ist nur faktisch, nicht aber rechtlich plural (§ 26 BGB / § 76 II 1 AktG[32]), sowie rechtlich (§ 28 BGB / § 77 I 1 AktG) und faktisch kollektiv.

Der pathologisch *zersplitterte Spartenvorstand* der Aktiengesellschaft wäre faktisch, aber nicht rechtlich plural (§ 76 II 1 AktG) und rechtlich (§ 77 I AktG) aber nicht faktisch kollektiv.

Der ordnungsgemäß besetzte *Aufsichtsrat* ist rechtlich (§ 95 S. 1 AktG) wie faktisch plural, und rechtlich (§ 108 I AktG) wie faktisch kollektiv.

Der *mit einer Person besetzte Aufsichtsrat* (§ 104 I AktG) ist rechtlich (§ 95 S. 1 AktG) aber nicht faktisch plural, sowie rechtlich (§ 108 I AktG) aber nicht faktisch kollektiv.

Die hier vorgeschlagene Differenzierung ermöglicht also eine genauere Beschreibung der Organstrukturen, die auch pathologische Fälle analytisch erfassbar macht. Statt also in einem bestimmten Fall festzustellen, ob ein Organ „kollegial" oder „nicht kollegial" ist, oder ob das Kollegialprinzip „gilt" oder „nicht gilt", bedarf es einer genauen Differenzierung, um die praktisch denkbaren Fälle in ihren Besonderheiten sachgerecht erfassen zu können.[33] Nur so lassen sich Missverständnisse vermeiden, wie sie in der bisherigen Literatur unweigerlich auftraten, wann immer ein Organ ohne nähere Erläuterung als „kollegial" bezeichnet wurde.

2. Systematik: Kollegien im Verwaltungs- und Justizorganisationsrecht

Als Prüfstein für die bisherige Konkretisierung des gesellschaftsrechtlichen Kollegialprinzips empfiehlt sich ein Blick in andere einschlägige Rechtsgebiete. Vor allem drei Rechtsgebiete haben bisher ein „Kollegialprinzip" herausgebildet, kultiviert und mehrfach monographisch aufgear-

---

[31] Als „rechtliche Anforderung" werden hier nur gesetzliche verstanden, nicht solche kraft Gesellschaftssatzung, denn es geht ja gerade um Anforderungen *an die* Satzung.

[32] Die dispositive Vorschrift des § 76 II 2 AktG ist keine rechtliche Verpflichtung, weil sie die Entscheidung für oder gegen Organpluralität gerade der Gesellschaft anheimstellt.

[33] Darüber hinaus ließe sich noch nach der Art der Organstellung differenzieren (z.B. Selbst-, Fremd- oder faktische Organschaft), aber das führt hier nicht weiter.

beitet: jene der Verwaltungs-,[34] Gerichts-[35] und Staatsorganisation.[36] Im folgenden untersuche ich, welches Begriffsverständnis diese Disziplinen entwickelt haben, und wie es mit der eben entwickelten Deutung vereinbar ist.

Als Kollegial*gremium* bezeichnet das Verwaltungsorganisationsrecht ein „aus mindestens drei Mitgliedern bestehendes, durch Rechtsnorm mit Beschlußkompetenzen ausgestattetes Gremium".[37] Ganz ähnlich wird der Begriff des Kollegial*organs* definiert als „ein Organ, das aus mehreren gleichberechtigt zusammenwirkenden und in einer rechtlichen Einheit organisierten Personen besteht."[38] Wenngleich diese Definitionen etwas voneinander abweichen, sind damit wohl keine Unterschiede in der Sache verbunden, wie die zitierten Werke an anderer Stelle verdeutlichen.[39] Nur die Voraussetzung gleichberechtigter Organmitglieder wird unterschiedlich streng gefasst und in ihrem engsten Verständnis teilweise für verzichtbar gehalten.[40] Der gemeinsame Nenner des verwaltungsorganisatorischen Verständnisses ist demnach das Kollegialorgan als Forum der *Beschlusstätigkeit mindestens dreier, meist gleichberechtigter, Entscheidungsträger*. Das entspricht der oben herausgearbeiteten Definition des gesellschaftsrechtlichen Kollegialprinzips (in seiner engen Ausprägung als Organkollektivität), mit einer Ausnahme: Das Verwaltungsorganisationsrecht setzt nicht allgemein „mehrere", sondern konkret mindestens drei Organmitglieder voraus. Wenngleich diese Voraussetzung „von geringerer praktischer Bedeutung" ist, wird ihr doch „von zahlreichen Autoren Aufmerksamkeit zuteil"[41] – auch im gesellschaftsrechtlichen Schrifttum.[42] Daher erscheint eine nähere Auseinandersetzung mit der Voraussetzung der Dreigliedrigkeit geboten.

---

[34] Vgl. oben Fn. 4 sowie den eingangs zitierten § 2 I BBremBG und § 88 VwVfG; Monographien von *Dagtoglou*, Kollegialorgane 1960, *Groß*, Kollegialprinzip 1999 und (aus Schweizer Sicht) *Ueberwasser*, Kollegialprinzip 1989.

[35] Vgl. oben Fn. 5; Monographien von *Guillard*, Kollegialgericht 1989 und – leider nur deskriptiv-rechtshistorisch – *Tüxen*, Kollegialprinzip 2009; vgl. auch *Rieß* in: Jung u.a., FS Müller 2008, 599 und *Jung* in: Hilgendorf/Rudolf, FS Heinz 2012, 883.

[36] *Schneider*, Beschlussfähigkeit 2000; *Thiele*, Entscheidungsfindung 2008, 57 ff. (Kap. 2) und 131 ff. (Kap. 3).

[37] *Groß*, Kollegialprinzip 1999, 46.

[38] *Dagtoglou*, Kollegialorgane 1960, 44; *Thiele*, Entscheidungsfindung 2008, 135 bei Fn. 35.

[39] Auch *Groß*, Kollegialprinzip 1999 erwähnt auf 49 f. die in seiner Definition scheinbar fehlende Gleichberechtigung der Gremienmitglieder; *Dagtoglou*, Kollegialorgane 1960 setzt auf 31 ff. die in seiner Definition nicht angesprochene Beschlusskompetenz und Dreigliedrigkeit voraus; auch *Ueberwasser*, Kollegialprinzip 1989, 1 weicht wohl sachlich nicht ab, wenn er kollegiales Handeln an der Führungslosigkeit festmacht statt an der Beschlussförmigkeit.

[40] Bspw. *Schneider*, Beschlussfähigkeit 2000, 22 f., u.a. mit Blick auf § 197 GVG und *Thiele*, Entscheidungsfindung 2008, 139, 156 mit Blick auf Sitz- und Stimmengewichtungen im Staatsrecht (vgl. 114 f.).

[41] *Thiele*, Entscheidungsfindung 2008, 138 m.w.N.

Die Mindestzahl von drei Mitgliedern wird zunächst damit begründet, dass schon im römischen Recht drei Mitglieder für ein „Kollegium" erforderlich gewesen seien.[43] Dieses traditionalistische Argument ist für die Rechtsauslegung freilich kaum ergiebig, solange nicht die konkreten Umstände des römischen Rechts und die sachlichen Erwägungen reflektiert werden, die der Dreierbesetzung zugrunde lagen. Viel spricht dafür, dass sie maßgeblich auf eine in den Religionen des Altertums vorherrschende Vorstellung von der **3** als „heiliger Zahl" zurückgeht und einer säkularen Rechtswissenschaft daher kaum noch Orientierung bieten kann.[44]

Nach wie vor bedenkenswert dagegen ist das sachliche Argument, dass erst ab drei Mitgliedern Mehrheitsbeschlüsse – als „eines der charakteristischen Merkmale von Kollegialgremien" und Voraussetzung für „eine plurale Struktur" – möglich seien.[45] Auch der Bundesgerichtshof meinte mit Blick auf Aufsichtsratsausschüsse, dass zweigliedrige Gremien zu „echten, im Sinne des Gesetzes liegenden kollegialen Entscheidungen" unfähig seien,[46] und berief sich zur Begründung auf die Gefahr „daß die beiden Mitglieder sich nicht einigen können und dann zum Nachteil des Unternehmens Entscheidungen unterbleiben, oder daß eines von ihnen, um überhaupt eine Entscheidung zu ermöglichen, schließlich nachgibt, ohne von den Argumenten des anderen überzeugt zu sein."[47] Denkt man diese Argumentation allerdings konsequent fort, wären auch Organe mit gerader Mitgliederzahl oder solche mit der Möglichkeit zur Stimmenthaltung nicht zu „echten" kollegialen Entscheidungen fähig, weil nie auszuschließen wäre, dass die Organmitglieder ihr Abstimmungsverhalten von der Angst vor Pattsituationen leiten lassen. Dabei belegen Erfahrungen aus der Justizorganisation, dass auch in zweigliedrigen oder in sonstiger gerader Zahl besetzten Spruchkörpern kaum „Dauerpatts" zu beobachten sind.[48] Auch in

---

[42] *Thamm*, Verfassung 2008, 109 Fn. 292, 297 m. Verw. auf *Groß* (Fn. 37); auch *Böckli*, Schweizer AktR 2009, 1732 springt in einem Satz unvermittelt von „mehrgliedrig sein" zu „mindestens drei Personen".

[43] Vgl. nur *Thiele*, Entscheidungsfindung 2008, 138 m.w.N.; *Groß*, Kollegialprinzip 1999, 47 m.w.N.

[44] So *Weber*, ZRP 1997, 134, 135; *Spoerri*, Bus Lawyer 1963, 305: "The potency of the magic number has survived into our day: the spell is only beginning to be broken."

[45] *Thiele*, Entscheidungsfindung 2008, 138 m.w.N.; *Groß*, Kollegialprinzip 1999, 47 m.w.N.; früher schon *Dagtoglou*, Kollegialorgane 1960, 32 f.; a.A. *Lorse*, ZBR 2003, 185, 186; genau umgekehrt *Roth Pellanda*, Organisation 2007, 212 (Rn. 435): Das dispositive Stichentscheidsrecht des Verwaltungsratsvorsitzenden (in der Schweiz Art. 713 I 2 OR, in Deutschland vgl. § 77 I 2 AktG) erfordere eine *gerade* Mitgliederzahl des Gremiums, widrigenfalls „das Institut des Stichentscheides toter Buchstabe bleiben würde."

[46] BGHZ 65, 190, 192; zust. zit. von *Thamm*, Verfassung 2008, 109 Fn. 296.

[47] BGHZ 65, 190, 192.

[48] *Weber*, ZRP 1997, 134, 135 mit Bsp. aus Deutschland, Frankreich und England

der GmbH gehört es zu den gesellschaftsvertraglichen Regelgestaltungen, dass zwei Geschäftsführer gemeinsam tätig werden können.[49] In der Aktiengesellschaft bildet die Gesamtgeschäftsführung *aller* Vorstandsmitgliesogar den gesetzlichen Regelfall (§ 77 I 1 AktG), so dass *jedem* Vorstandsmitglied ein Vetorecht gegen alle anderen zukommt.[50] Wer also den Bundesgerichtshof so versteht, dass Vetorechte mit kollegialer Entscheidungsfindung unvereinbar seien,[51] müsste schon in der einstimmigen Gesamtgeschäftsführung nicht „die reinste Form" des Kollegialprinzips,[52] sondern geradezu dessen Widerspruch sehen.

Allein aus der Tatsache, dass zweigliedrige Organe sich unter Umständen nicht einigen können, folgt mithin nicht, dass sie generell nicht kollegial entscheiden könnten. Der Bundesgerichtshof argumentiert zwar weiter, dass in „einem aus nur zwei Mitgliedern gebildeten Gremium nicht in gleichem Maße" anzunehmen sei, „daß möglichst alle für eine Entscheidung maßgebenden sachlichen Gesichtspunkte zur Sprache kommen und Einseitigkeiten vermieden werden";[53] mit Tatsachen belegt er diese Behauptung jedoch nicht. Ihr ist unter Hinweis auf § 76 II 2 AktG (der ja zweigliedrige Vorstände ausdrücklich ermöglicht) entgegenzuhalten: „Es gibt keinen allgemeinen Grundsatz, dass nur ein mindestens dreiköpfiges Gremium sachgerechte Entscheidungen trifft."[54]

Ein letzter Ansatzpunkt dafür, dass Kollegialorgane mindestens dreigliedrig sein müssen, könnte sich allenfalls noch aus § 108 I AktG ergeben, wonach der Aufsichtsrat „durch Beschluss" entscheiden und deshalb womöglich Mehrheitsbeschlüsse fassen können muss. Diese Norm gilt allerdings nur für den Aufsichtsrat, der nach § 95 S. 1 AktG ohnehin mindestens drei Mitglieder haben muss. Im Allgemeinen ist die Mehrheitsbeschlussfähigkeit dagegen kein begriffsnotwendiges Merkmal einer kollegialen Entscheidung, denn Kollegialität kann ebenso gut nach dem Einstimmigkeitsprinzip organisiert sein.[55] Das ist für den Vorstand der Aktiengesellschaft nach § 77 I 1 AktG sogar der gesetzliche Regelfall.

Demnach lässt sich festhalten: Vielleicht sind Zweiergremien „kein »Kollegium« im eigentlichen Sinne", doch auch „[f]ür den zweiköpfigen

---

[49] *Stephan* in: Hoffmann-Becking/Rawert, Formularbuch 2010, IX.1 Anm. 17 Abs. 2.

[50] So ausdr. *Seibert* in: Lutter/Schmidt, AktG 2010, § 77 Rn. 14, vgl. auch Rn. 5, 8; *Hüffer*, AktG 2012, § 77 Rn. 6, 12; *Spindler* in: Goette/Habersack, MK-AktG 2014, § 77 Rn. 10, 16.

[51] So offenbar *Thamm*, Verfassung 2008, 109 bei Fn. 293.

[52] *Bleicher/Leberl/Paul*, Unternehmensverfassung 1989, 30.

[53] BGHZ 65, 190, 192; wortgleich *Thamm*, Verfassung 2008, 109.

[54] *Werner*, AG 1976, 45, 46; ebenso Begr. Art. 1 Nr. 3 Justizbeschleunigungsgesetz, BT-Drs. 15/1491 v. 28.8.2003, 17: „Die Vorteile des Kollegialprinzips […] sind auch bei einer Zweierbesetzung des Kollegiums im Wesentlichen noch gegeben".

[55] So schon *Weber*, Wirtschaft 1922, 159; vgl. auch *Ueberwasser*, Kollegialprinzip 1989, 8.

Vorstand gilt das Kollegialprinzip".[56] Das gesellschaftsrechtliche Kollegialprinzip setzt also – anders vielleicht als das verwaltungsorganisatorische – keine allgemeine Untergrenze bei drei Kollegen.

Eine zweite Frage ist, ob das Kollegialprinzip eine Obergrenze kennt. Die bisher entwickelte Definition schließt es nicht aus, dass Kollegialorgane mit mehreren hundert Mitgliedern besetzt sind, und dementsprechend gelten im Staatsorganisationsrecht sogar gesetzgebende Versammlungen als Kollegialorgane.[57] Ganz ähnlich lassen sich im Gesellschaftsrecht auch die Hauptversammlung der Aktiengesellschaft und die Mitgliederversammlung des Vereins als Kollegialorgane kraft rechtlicher (§ 56 BGB) bzw. faktischer Organpluralität (§ 2 AktG) und rechtlicher Organkollektivität (§ 32 I 1 BGB / § 118 I 1 AktG) beschreiben. Dieses Verständnis würde sich allerdings so weit vom etablierten gesellschaftsrechtlichen Sprachgebrauch entfernen,[58] dass eine pragmatische Trennlinie erforderlich ist. Als Kriterium für eine solche Trennlinie wurde im Staatsorganisationsrecht die Differenzierung zwischen „plenaren" und „nicht plenaren" Kollegialorganen vorgeschlagen, die sich danach richten soll, ob die Mitgliederzahl des Organs beschränkt oder unbeschränkt ist.[59] Das kann aber schon im Staatsorganisationsrecht nicht überzeugen – kennt es doch sowohl plenare Organe mit beschränkter Mitgliederzahl (§ 1 I 1 BWahlG) als auch nicht plenare Organe mit unbeschränkter Mitgliederzahl (Artt. 62, 64 I GG) – und ist im Gesellschaftsrecht nicht minder problematisch.

Mit Blick auf die bevorstehende Untersuchung bietet sich eher eine Abgrenzung danach an, ob das betreffende Organ zu einer Beratung in der Lage ist, in der die Organmitglieder wechselseitig zu Wort kommen und sich unmittelbar aufeinander beziehen können. Kollegialorgane, in denen das nicht der Fall ist und die ihre Beschlussfassung vorrangig durch Frontalvorträge, Reden sowie förmliche Gegen- und Änderungsanträge vorbereiten, fallen aus der vorliegenden Betrachtung heraus.[60]

---

[56] *Priester*, AG 1984, 253, 255; auch *Lorse*, ZBR 2003, 185, 186 setzt das voraus, wenn er das Dreierdogma unter Verweis auf das Gesellschaftsrecht kritisiert; a.A. *Thamm*, Verfassung 2008, 110 bei Fn. 299.

[57] *Thiele*, Entscheidungsfindung 2008, 139; vgl. auch *Haymann* in: Tatarin-Tarnheyden, FS Stammler 1926, 395, 429 ff.

[58] Mir ist kein Fall bekannt, in dem diese Versammlungen überhaupt je als „kollegial" bezeichnet worden wären.

[59] *Thiele*, Entscheidungsfindung 2008, 152.

[60] Nochmals zu betonen ist allerdings, dass diese Schwerpunktsetzung einer gewissen pragmatischen Willkür geschuldet ist. Das Verhältnis zwischen dem Kollegialprinzip und Gesellschafterversammlungen bedürfte einer vertieften Untersuchung, die hier allerdings nicht zu leisten ist.

3. Rechtsvergleich: Zwanzig Gesellschaftsrechte im Kontrast

Nach der oben gewonnenen und am Sprachgebrauch des Verwaltungsrechts näher eingegrenzten Begriffskonkretisierung untersuche ich nun die rechtliche Bedeutung und internationale Verbreitung des Kollegialprinzips, bevor ich seinen Sinn und Zweck hinterfrage.

Nach gegenwärtigem Stand der funktionalen Rechtsvergleichung auf dem Gebiet des Gesellschaftsrechts gehört es zu den fünf allgemeinen Strukturprinzipien der Kapitalgesellschaft (*corporation*), dass sie durch ein besonderes Organ (*board*) verwaltet wird.[61] Als eines der vier kennzeichnenden Merkmale dieses Organs gilt, dass es „gewöhnlich mehrere Mitglieder hat".[62] Damit ist allerdings noch nicht gesagt, ob diese plurale Besetzung nur tatsächlich praktiziert wird oder auch rechtlich vorgeschrieben ist, und ob sie mit einer gemeinschaftlichen Entscheidungsfindung im Sinne der oben definierten Organkollektivität einhergeht. Um diese Fragen zu beantworten, bedarf es eines genaueren Blicks in einzelne Rechtsordnungen.[63] Dazu untersuche ich nun das Gesellschaftsrecht 19 ausgewählter Länder[64] sowie die paneuropäische Rechtsform *Societas Europaea* (SE). Die ausgewählten Länder bzw. ihre Rechtsordnungen stehen stellvertretend erstens für fünf der sechs Kontinente,[65] zweitens für alle nach herkömmlicher Lesart abgegrenzten Rechtsfamilien[66] und drittens für die weltgrößten Wirtschaftsnationen.[67]

Für jede Rechtsordnung wurde anhand des neueren deutsch- und englischsprachigen Schrifttums erhoben, welche gesellschaftsrechtlichen Vorgaben zur pluralen Organbesetzung und zur kollektiven Entscheidungsfindung existieren – und welche Zwecke der jeweilige Gesetzgeber damit ver-

---

[61] *Kraakman/Armour u.a.*, Anatomy 2009, 5 ff., 12.

[62] *Kraakman/Armour u.a.*, Anatomy 2009, 14: "the board ordinarily has multiple members. […] However, there are exceptions."

[63] Etwa stellt *Jung* in: Hilgendorf/Rudolf, FS Heinz 2012, 883, 891 „unterschiedliche Einschätzungen des Kollegialprinzips im internationalen Vergleich" fest, die der „rechtskulturellen Vertiefung vermittels einer sozialpsychologisch unterfütterten Rechtsvergleichung" bedürfen; zur sozialpsychologischen Unterfütterung nachher B.VII.

[64] Australien, Brasilien, China, Deutschland, Finnland, Frankreich, Indien, Italien, Japan, Kanada, Österreich, Polen, Russland, Schweden, Schweiz, Spanien, Südkorea, USA und Vereinigtes Königreich.

[65] Afrikanische Länder fehlen in der vorigen Fn.

[66] Nach der Klassifizierung von *Zweigert/Kötz*, Rechtsvergleichung 1996, 62 ff. (§§ 5–21) sind in Fn. 64 der romanische Rechtskreis mit vier, der deutsche mit drei, der anglo-amerikanische mit fünf, der nordische mit zwei und der fernöstliche mit drei Rechtsordnungen vertreten, hinzu kommen die Übergangssysteme Polens und Russlands; krit. *Husa*, Rev Int Droit Comp 2004, 11 mit Alternativklassifizierung.

[67] Die Volkswirtschaften in Fn. 64 belegen in *World Bank*, GNI 2010 die Ränge 1–13, 15, 19, 21, 22, 25 und 34.

folgte, worauf ich unter II. gesondert eingehe. Diese Erhebung beschränkt sich weitgehend auf die Kapitalgesellschaften (darunter oft eine geschlossene Rechtsform wie die GmbH und eine kapitalmarktoffene wie die AG) wurde aber für die Rechtssysteme des deutschen Sprachraums gezielt um Vereine als Grundform der privaten Körperschaft ergänzt. Teilweise wurden auch informelle Regulierungsschichten (wie Börsenzulassungsregeln und Corporate Governance Kodizes) gesichtet. Die Rechtslage stellt sich in den zwanzig Rechtsordnungen wie folgt dar:

Im *australischen* Kapitalgesellschaftsrecht mussten Gesellschaften früher grundsätzlich ein monistisches Verwaltungsorgan aus drei Direktoren haben; erst das Gesellschaftsrechtsvereinfachungsgesetz von 1995 befreite geschlossene Gesellschaften von dieser Mindestzahl.[68] Seither müssen geschlossene Gesellschaften nur noch einen Direktor, offene dagegen mindestens drei haben (§ 201A Corporations Act).[69] Implizit geht das Gesetz dabei davon aus, dass die Direktoren in kollegialer Beratung entscheiden.[70]

Das *brasilianische* Gesellschaftsrecht kennt sowohl eine der GmbH vergleichbare Rechtsform, die seit einer Novelle von 2003 im Bürgerlichen Gesetzbuch geregelt ist,[71] als auch „geschlossene" und „offene" Aktiengesellschaften nach dem Gesetz Nr. 6.404 von 1976. Während das GmbH-Äquivalent selbstorganschaftlich geleitet wird, wenn nicht ein oder mehrere Geschäftsführer bestellt sind,[72] muss die geschlossene Aktiengesellschaft einen aus mindestens zwei Direktoren bestehenden „Vorstand" (Art. 143 LSA) und einen aus drei bis fünf Mitgliedern bestehenden „Überwachungsrat" haben (Art. 161 § 1 LSA), die offene Aktiengesellschaft darüber hinaus noch einen Verwaltungsrat, der als „Organ gemeinschaftlicher Beschlussfassung" aus mindestens drei Mitgliedern besteht (Art. 138 § 2, Art. 140 LSA).[73]

Das *chinesische* Kapitalgesellschaftsrecht (für Hongkong gelten Sonderregeln[74]) verfügt über zwei Rechtsformen, die beide dualistisch aufgebaut

---

[68] *Austin/Ford/Ramsay*, Company Directors 2005, 59 f.

[69] *Tomasic/Bottomley/McQueen*, Corporations Law 2002, 263; *Austin/Ford/Ramsay*, Company Directors 2005, 59; *Fritzemeyer* in: Süß/Wachter, Hdb GmbH 2011, 438.

[70] *Austin/Ford/Ramsay*, Company Directors 2005, 114: "It is an assumption of the Corporations Act that the board will arrive at a decision at a meeting of directors."; vgl. schon *Tomasic/Bottomley/McQueen*, Corporations Law 2002, 264: "an assumption that directors will exercise this power of management as a group […] meet together to conduct their business and to make decisions." m.w.N. aus der Rspr. in Fn. 6 und 7.

[71] *Curschmann/Jolowicz*, GmbHR 2003, 1185, 1190.

[72] *Curschmann* in: Süß/Wachter, Hdb GmbH 2011, 502 f., 517 f. m.Verw. auf Artt. 1060, 1053 i.V.m. 1013 NCC.

[73] *Florence*, Brasilianisches GftsR 2005, 10 f. mit deutscher Gesetzesfassung (29 ff.).

[74] Z.B. *Gu*, Chinese Law 2006, 134: "not less than 2 directors are required (Section 153 of the Hong Kong Companies Ordinance)."

und im „Gesellschaftsgesetz“ (Gōngsīfǎ) von 1993 geregelt sind, das zum 1.1.2006 neu gefasst wurde.[75] Der „Vorstand“ muss in der chinesischen GmbH drei bis 13 Mitglieder umfassen (§ 45 Gōngsīfǎ), in der chinesischen AG fünf bis 19 Mitglieder (§ 109 Gōngsīfǎ), der „Aufsichtsrat“ muss in beiden Gesellschaftsformen jeweils mindestens drei Mitglieder haben (§§ 52, 118 Gōngsīfǎ).[76] Trotz dieser Bedeutung mehrgliedriger Organe ist deren kollektive Entscheidungsfindung „nicht explizit als Regelfall vorgeschrieben“, sondern lässt sich aus dem Gesetz nur andeutungsweise ablesen, „dass der Vorstand der chinesischen Aktiengesellschaft kollegial zu organisieren ist [… also] die Geschäfte der Gesellschaft durch Beschlussfassung auf der Vorstandssitzung gemeinschaftlich führen soll.“[77] Jedenfalls sind aber GmbH „mit verhältnismäßig wenig Gesellschaftern oder verhältnismäßig kleinem Umfang“ von der Pflicht ausgenommen, überhaupt einen Mehrpersonenvorstand und einen Aufsichtsrat zu bilden (§§ 51, 52 Gōngsīfǎ).[78] Ebenso fakultativ in allen GmbH – dafür zwingend in den Aktiengesellschaften – ist die Bestellung eines „Managers“ neben dem Vorstand (§§ 50 I, 119 I 1 Gōngsīfǎ), der als Überbleibsel aus der früheren Betriebsleitung durch staatliche Kader[79] allerdings kein Organ darstellt, sondern nur Durchführungsgehilfe für die Entscheidungen des Vorstands ist.[80]

Im *deutschen* Gesellschaftsrecht ist allein der Aufsichtsrat einer Aktiengesellschaft zwingend mehrgliedrig – nämlich mit drei bis 21 Mitgliedern – zu besetzen (§§ 95 S. 1, 108 II 3 AktG). Die Vorstände großer Aktiengesellschaften (3 Mio. Euro Grundkapital) benötigen lediglich dispositiv mindestens zwei Mitglieder (§ 76 II 2 AktG), wenngleich das auf dem Umweg der Mitbestimmung für alle Kapitalgesellschaften einer gewissen Größe (2.000 Mitarbeiter) zwingend gilt (§ 33 I 1 MitbestG).[81] Im Übrigen wird das „Vieraugenprinzip“ nur „als Best Practice“ mit Erklärungspflicht

---

[75] Englische Fassung abgedr. in *CCH Asia*, China II 2005, 150.011 ff.; deutsche Fassung online in *Münzel*, Chinas Recht 2009.

[76] *Kroymann*, China 2009, 297, 301; *Koh Soon Kwang*, Company Law 2010, 118 (Rn. 5.66); *Bu*, Einführung 2009, 177 (Rn. 37, 40); *Scheil* in: Süß/Wachter, Hdb GmbH 2011, 560.

[77] *Hu*, Rechtsfragen 2006, 117 bzw. 116.

[78] *Bu*, Einführung 2009, 174 (Rn. 29); krit. *Jiang*, GmbH in China 2011, 116: „Der konkrete Maßstab für ‚verhältnismäßig wenige Gesellschafter‘ oder ‚verhältnismäßig kleiner Umfang‘ wird nicht von dem GesG vorgegeben, was bedeutet, dass die GmbH fast völlig nach eigenem Bedürfnis entscheiden kann, ob sie statt eines Vorstands nur einen geschäftsführenden Vorsteher bestellt.“

[79] *Jiang*, GmbH in China 2011, 119 f.

[80] *Hu*, Rechtsfragen 2006, 119 f.; ähnl. *Gu*, Chinese Law 2006, 175: “a manager is only an assistant organ for the [Board of Directors], so is not regarded as an important organ”.

[81] Auch rechtstatsächlich haben „Mehrpersonenvorstände […] sich mehr und mehr zur Regel entwickelt“, so *Emde* in: Burgard u.a., FS Schneider 2011, 295, 297.

nach § 161 AktG empfohlen (Ziff. 4.2.1 DCGK)[82] oder sprachlich nahegelegt (§ 35 I 1 GmbHG: „die Geschäftsführer"). Wo sich das Gesetz auch solch schwacher Fürsprache für die Organpluralität enthält, ermöglicht es zumindest ausdrücklich die Besetzung des Organs mit mehreren Mitgliedern – so die §§ 26 II, 28 BGB für den Vereins- und § 76 II 1 AktG für den AG-Vorstand. Wenngleich also die Organpluralität in Deutschland sehr zurückhaltend angeordnet ist, gilt die Organkollektivität dennoch durchgehend: Im aktienrechtlichen Vorstands-[83] und Aufsichtsratshandeln[84] ebenso wie im Zusammenwirken mehrerer Geschäftsführer im GmbH-Recht[85] sowie den Vorstandsbeschlüssen im Vereinsrecht.[86] Das wird aus ganz verschiedenen Vorschriften hergeleitet,[87] wobei erneut die mangelnde Konsistenz im Sprachgebrauch zum Tragen kommt (oben bei Fn. 19–21): wer etwa Kollegialprinzip und Gesamtgeschäftsführung gleichsetzt, verortet das Prinzip natürlich anders als jemand, der es synonym mit der Gesamtvertretung verwendet. Angesichts seiner rechtsformübergreifenden Anwendbarkeit erscheint es am überzeugendsten, die Organkollektivität als

[82] *Ringleb* in: Ringleb/Kremer u.a., DCGK 2014, Rn. 646; vgl. auch *Spindler* in: Goette/Habersack, MK-AktG 2014, § 76 Rn. 97: „Gewöhnlich, insbesondere bei allen größeren Gesellschaften, wird die Bildung eines mehrköpfigen Vorstands erforderlich sein."; noch konkreter *IK Berlin*, DB 2000, 1573, 1576: „Der Vorstand soll im Normalfall mindestens drei und höchstens neun Mitglieder haben."

[83] *Wettich*, Vorstandsorganisation 2008, 12; *Martens* in: Goerdeler u.a., FS Fleck 1988, 191; Schon vor 1937 galt unter § 231 HGB nichts anderes: *Staub*, HGB 1933, § 231 Anm. 8 m.w.N.; *Brodmann*, AktR 1928, § 231 Anm. 2 (S. 217).

[84] RGZ 93, 338, 340; Begr. § 108 AktG-RegE 1965, BT-Drs. 4/171 in: *Kropff*, AktG 1965, 151: „Aufsichtsrat [.] als Kollegium"; *Habersack* in: Goette/Habersack, MK-AktG 2014, vor § 95 Rn. 14; *Lutter*, ZHR 1981, 224, 250: „Der Aufsichtsrat ist Kollegialorgan."

[85] RGZ 91, 72, 77; RGZ 98, 98, 100; BGHZ 133, 370, 376 f.; *Schneider* in: Lutter u.a., FS GmbH 1992, 473, 481 ff.; *Fleischer* in: Fleischer/Goette, MK-GmbHG 2011, § 43 Rn. 112 m.w.N. in Fn. 1; *Haas/Ziemons* in: Michalski, GmbHG 2010, § 43 Rn. 153; uneindeutig *Thamm*, Verfassung 2008, 88: „In der GmbH können mehrere Geschäftsführer bestellt werden, die dann ebenfalls ein Kollegium bilden können."

[86] *Schießl/Küpperfahrenberg*, DStR 2006, 445, 446; *Reuter* in: Säcker, MK-BGB 2012, § 27 Rn. 44 a.E. m.w.N.; *Waldner/Wörle-Himmel* in: Sauter/Schweyer/Waldner, Verein 2010, Rn. 277A bezweifelt die Gesamtverantwortung der Vereinsvorstände nur soweit sie gegenseitige Überwachungspflichten umfassen soll; diff. *Heermann* in: Crezelius u.a., FS Röhricht 2005, 1191, 1193, 1197.

[87] Mögliche Anknüpfungspunkte im Aktienrecht z.B. § 76 I AktG (so *Wettich*, Vorstandsorganisation 2008, 35), ggf. i.V.m. der Organstellung der Geschäftsleiter (so *Fleischer*, NZG 2003, 449, 450), oder § 77 I AktG (so *Hanau*, ZGR 1983, 346, 369; *Oesterle*, zfo 2003, 199, 200; *Dauner-Lieb* in: Henssler/Strohn, GftsR 2014, § 77 Rn. 1; *von Werder* in: Ringleb/Kremer u.a., DCGK 2014, Rn. 96) oder die „gesamten §§ 76-78" (so *Beckert*, Personalisierte Leitung 2009, 37; ähnl. *Dauner-Lieb* in: Henssler/Strohn, GftsR 2014, § 78 Rn. 1), im GmbH-Recht z.B. als Verallgemeinerung von § 35 II 2 GmbHG (so wohl *Haas/Ziemons* in: Michalski, GmbHG 2010, § 43 Rn. 153).

Ausdruck eines „organisationsrechtlichen Vorverständnisses“ zu qualifizieren, das „der gesetzlichen Regelung der Kollegialorgane im Vereinsrecht, GmbH-Recht und Aktienrecht unausgesprochen zugrunde liegt.“[88]

Das *finnische* Kapitalgesellschaftsrecht kennt keine der GmbH vergleichbare Rechtsform,[89] unterscheidet aber öffentliche und private Aktiengesellschaften.[90] Bis zur Novelle 2006, die das Aktienrecht im Interesse kleiner und mittelständischer Unternehmen flexibilisierte,[91] mussten öffentliche Aktiengesellschaften, sowie private mit einem Grundkapital in entsprechender Höhe (80.000 Euro), den Vorstand mindestens dreigliedrig ausgestalten (OYL 8:1.1 a.F.).[92] Seit 2006 genügt schon ein Vorstandsmitglied, wenn ein Vertreter bestellt ist, und darf der Vorstand höchstens fünf Mitglieder haben (OYL 6:8.1).[93] Der Aufsichtsrat, der neben dem dreigliedrigen Vorstand bis 2006 eingerichtet werden musste (OYL 8:11.1 a.F.) aber ohnehin selten anzutreffen war,[94] ist nunmehr fakultativ (OYL 6:1.1), muss aber weiterhin mindestens drei Mitglieder aufweisen (OYL 6:23).[95]

Im *französischen* Kapitalgesellschaftsrecht gibt es eine auf 100 Gesellschafter beschränkte GmbH (L.223-3 CCom), in der hinsichtlich der Organpluralität Gestaltungsfreiheit herrscht,[96] wenngleich mehrere Geschäftsführer zwingend einzelvertretungsbefugt sind.[97] Daneben existieren vier Gestaltungsvarianten der Aktiengesellschaft: die traditionelle monistische, die „weit weniger beliebte“[98] dualistische, die gemischte und seit 1994

---

[88] *Hoffmann-Becking*, ZGR 1998, 497, 506 f.; *Hoffmann-Becking*, NZG 2003, 745, 746 f. (mit guten Argumenten ausdrücklich gegen die Verortung etwa in § 76 I oder § 77 I AktG); zust. *Wettich*, Vorstandsorganisation 2008, 40.

[89] *van Setten* in: Ars Legis e.V., Kapitalgesellschaften 2007, 97; *Nicolaysen*, RIW 2005, 884, 885: in Skandinavien haben nur Dänemark und Norwegen GmbH-Äquivalente eingeführt; vgl. aber *Kocher*, Finnland 2003, 33: „Die private AG entspricht zwar nicht rechtlich, wohl aber in ihrer wirtschaftlichen Verwendung weitgehend der deutschen GmbH.“

[90] *Kocher*, Finnland 2003, 32, auf 33 mit dem Hinweis, dass öffentliche Aktiengesellschaften „fast nur für wirklich große Unternehmen benutzt“ würden.

[91] *Miettinen*, RIW 2006, 812, passim.

[92] Englische Fassung abgedr. bei *Snellman*, Companies Act 1984, 38; vgl. auch *Kocher*, Finnland 2003, 32.

[93] Englische Fassung (neu) unter www.finlex.fi/en/laki/kaannokset/2006/en20060624; vgl. auch *Faber* in: Süß/Wachter, Hdb GmbH 2011, 863.

[94] *Van Hulle/Gesell*, Corporate Law 2006, 130: “very few [.] are established”.

[95] Vgl. auch *Faber* in: Süß/Wachter, Hdb GmbH 2011, 865.

[96] *Feuerbach/Victor-Granzer* in: Ars Legis e.V., Kapitalgesellschaften 2007, 79; *Van Hulle/Gesell*, Corporate Law 2006, 151; Hopt/Leyens, ECFR 2004, 135, 156 ff.; vgl. früher schon Art. 49 des Gesetzes Nr. 66-537 über die Handelsgesellschaften, auf Deutsch abgedr. in *Gravenstein*, Französisches GftsR 1988, 65.

[97] *Sonnenberger/Autexier*, Einführung 2000, 193; *Karst* in: Süß/Wachter, Hdb GmbH 2011, 901.

[98] *Hübner/Constantinesco*, Einführung 1994, 222; laut *Arlt*, Französische AG 2006, 73 m.w.N. waren 25 % der börsennotierten, aber nur 3 % *aller* Aktiengesellschaften dualistisch

auch noch die vereinfachte.[99] In der monistischen Aktiengesellschaft ist ein Verwaltungsrat als „Mehrpersonengremium“ mit drei bis 18 Mitgliedern einzurichten (L.225-17 I CCom).[100] Er ist „seiner grundlegenden Konstitution nach ein kollegiales Gremium, das ausschließlich kollektive Entscheidungen treffen darf.“[101] Dem entspricht in der dualistischen Gestaltungsvariante das Direktorium als ein „dem Grundsatz nach kollegiales Gremium“[102] mit zwei bis fünf Mitgliedern, ausnahmsweise weniger in Gesellschaften mit einem Grundkapital bis 150.000 Euro, oder mehr (bis zu sieben Direktoren) in börsennotierten Gesellschaften (L.225-58 CCom).[103] Das Direktorium wird von einem Aufsichtsrat mit drei bis 18 Mitgliedern überwacht (L.225-69 I CCom), der „seiner Konzeption nach in wesentlichen Teilen dem Verwaltungsrat einer monistischen Aktiengesellschaft nachgebildet“ ist und ebenso wie jener „als kollegiales Organ […] seine Aufgaben und Funktionen im Wege kollektiver Entscheidungsfindung wahrnimmt.“[104] Um „umständliche, zeitintensive Geschäftsleitungsorgane“ zu vermeiden, wurde 1994 die vereinfachte Aktiengesellschaft eingeführt, die bei der Organbesetzung völlige Gestaltungsfreiheit genießt (L.227-5 CCom).[105]

Die *indischen* Kapitalgesellschaften haben ein monistisches Geschäftsführungsorgan, das spätestens seit 1913 mindestens zwei Direktoren aufweisen muss, wenn die Gesellschaft geschlossen ist, und mindestens drei Direktoren, wenn sie am Kapitalmarkt auftritt.[106] Diese Mindestzahlen wurden im Companies Act von 1956 in § 252 I, II niedergelegt[107] und in den späteren Reformen von 1965 und 2000 nicht wesentlich geändert.[108]

---

organisiert; ähnl. *Menjucq*, ZGR 2003, 679, 680: 20 % aller börsennotierten, 3 % insgesamt; *Sonnenberger/Autexier*, Einführung 2000, 195: 2,3 % aller Aktiengesellschaften 1997.

[99] *Arlt*, Französische AG 2006, 73 f.; Hopt/Leyens, ECFR 2004, 135, 156 ff.

[100] *Peifer*, Kapitalgesellschaften 2009, 351 (Rn. 596); *Feuerbach/Victor-Granzer* in: Ars Legis e.V., Kapitalgesellschaften 2007, 80; aktuelle englische Fassung online in *Vogel/Perochon*, C.com. 2006, 48 f.; vgl. früher schon Art. 89 des Gesetzes Nr. 66-537 (oben Fn. 96).

[101] *Peifer*, Kapitalgesellschaften 2009, 365 (Rn. 621); ähnl. schon *Wietek/Chomiac de Sas*, Die SAS 2003, 24: „hybrides Kollegialorgan“; *Menjucq*, ZGR 2003, 679, 681: „kollegiales Leitungsorgan“; *Arlt*, Französische AG 2006, 217: „Kollegialorgan“.

[102] *Peifer*, Kapitalgesellschaften 2009, 390 (Rn. 677); *Feuerbach/Victor-Granzer* in: Ars Legis e.V., Kapitalgesellschaften 2007, 80 f.: „Kollegialorgan“; ebenso *Arlt*, Französische AG 2006, 177; *Beckert*, Personalisierte Leitung 2009, 127 m.w.N. in Fn. 663.

[103] *Peifer*, Kapitalgesellschaften 2009, 390 f. (Rn. 678 f.).

[104] *Peifer*, Kapitalgesellschaften 2009, 400 (Rn. 697 f.); vgl. früher schon Art. 129 des Gesetzes Nr. 66-537 (oben Fn. 96).

[105] *Wietek/Chomiac de Sas*, Die SAS 2003, 13, 98: „Man könnte keine schlechtere Wahl treffen, als die einfache AG sofort wieder mit allen Arten von Kollegialorganen zu belasten.“

[106] *Swarup*, Companies Act 1966, § 252 Rn. 1.

[107] *Wegen*, GftsR in Indien 1997, 34; abgedr. bei *Swarup*, Companies Act 1966, 560.

[108] Näher *Ramaiya/Chandrachud*, Companies Act 2006, 2591.

Auch die jüngste Novelle von 2013 hat diese Mindestzahlen beibehalten, zugleich aber die Einpersonengesellschaft ermöglicht (§ 149 I a Companies Act 2013).[109] Die Höchstzahl beträgt seither 15 Direktoren (§ 149 I b Companies Act 2013). Diese führen die Geschäfte gemeinschaftlich.[110]

Das *italienische* Kapitalgesellschaftsrecht kennt zunächst ein Äquivalent zur GmbH, das im Grunde selbstorganschaftlich ohne zwingende Organpluralität verwaltet wird, aber ab bestimmten Buch- oder Bilanzwerten einen der Aktiengesellschaft nachgebildeten Überwachungsrat braucht (Art. 2477 IV c.c.).[111] Die Aktiengesellschaft darf ihrerseits seit 2004 zwischen jener traditionellen Struktur und ihren monistischen und dualistischen Alternativen wählen.[112] Sind in der traditionellen Struktur mehrere Verwalter bestellt, üben sie als Verwaltungsrat „die Geschäftsführung […] als Kollegialorgan aus. Willensbildung und Entscheidung erfolgen durch Beschluss."[113] Auch der Überwachungsrat, der aus drei oder fünf Mitgliedern sowie zwei Ersatzmitgliedern besteht (Art. 2397 I c.c.), „entscheidet als Kollegialorgan durch Beschlüsse".[114] In der dualistischen Struktur (Art. 2409-octies ff. c.c.) braucht der Vorstand mindestens zwei, der Aufsichtsrat mindestens drei Mitglieder,[115] allerdings „folgt die Praxis weiterhin fast einhellig dem traditionellen System – die neuen Vorschriften zur dualistischen und zur monistischen Verwaltungsstruktur sind weithin *‚lettera morta'*."[116]

Das *japanische* Kapitalgesellschaftsrecht verfügt neuerdings nur noch über die Aktiengesellschaft, nachdem das frühere GmbH-Äquivalent – das „kein Kollegialorgan" vorsah, sondern nur einen Geschäftsführer, § 25 jGmbHG a.F.[117] – durch das neue „Gesellschaftsgesetz" (Kaishahô) zum

---

[109] Aktuelle und vorige Fassungen in *Gov't of India*, Companies Act 2014.

[110] *Ramaiya/Chandrachud*, Companies Act 2006, 2596: "The directors must act as a body, under the authority of a meeting properly convened and without improper exclusion of any of the directors."

[111] *Gebhard* in: Ars Legis e.V., Kapitalgesellschaften 2007, 162 f., sowie 164 zu den genauen Schwellenwerten; *Van Hulle/Gesell*, Corporate Law 2006, 208; laut *Kindler*, ZEuP 2012, 72, 77 besteht eine Wahl zwischen „Buchprüfer" und „Prüferkollegium"; deutsche Fassung online in *Bauer/Eccher u.a.*, Italienisches Zivilgesetzbuch 2010.

[112] *Kindler*, Einführung 2008, 314; *Kindler*, ZEuP 2012, 72, 74 ff.; *Fasciani* in: Süß/Wachter, Hdb GmbH 2011, 953 f.; *Gebhard* in: Ars Legis e.V., Kapitalgesellschaften 2007, 159; *Van Hulle/Gesell*, Corporate Law 2006, 215.

[113] *Schnieders*, Italien 2009, 107; *Kindler*, ZEuP 2012, 72, 74.

[114] *Schnieders*, Italien 2009, 114.

[115] *Gebhard* in: Ars Legis e.V., Kapitalgesellschaften 2007, 165 f.; *Kindler*, ZEuP 2012, 72, 76.

[116] *Kindler*, ZEuP 2012, 72, 79 mit Fn. 41: Nur 476 Aktiengesellschaften hatten sich bis 2008 für eine der beiden Alternativstrukturen entschieden.

[117] *Kawamoto*, Japan 2004, 253 (Rn. 655 f.); *Menkhaus/Pawlowski* in: Süß/Wachter, Hdb GmbH 2011, 999.

1.5.2006 abgeschafft worden ist.[118] Japanische Aktiengesellschaften werden traditionell dualistisch geführt, durch einen Verwaltungsrat, der drei Mitglieder umfassen muss (Art. 327 I Nr. 1 Kaishahô, zuvor Art. 225 jHGB a.F.),[119] sowie einen Prüferrat gleicher Mindestgröße (331 IV, 335 III Kaishahô).[120] Seit der Reform gelten diese Mindestgrößen nur noch für „große Publikumsgesellschaften", deren Anteile frei handelbar und nicht vinkuliert sind.[121] Das Leitungsorgan kleiner geschlossener Aktiengesellschaften besteht dagegen „entweder aus einem oder mehreren Direktoren, die kein Kollegialorgan bilden, Art. 326 Abs. 1 *kaisha hō*, oder aus einem Kollegialorgan".[122] Große Aktiengesellschaften haben in der Praxis oft sehr große Verwaltungsräte, bisweilen mit über 70 Mitgliedern,[123] deshalb wurde die gesetzliche Kompetenzregelung früher meist durch Leitungsausschüsse ausgehöhlt, die ohne gesetzliche Grundlage eingerichtet wurden.[124] Seit 2002 können große Aktiengesellschaften auch offiziell ein monistisches Ausschusssystem einrichten, in dem der Verwaltungsrat mindestens drei Ausschüsse mit je drei Mitgliedern bilden muss.[125] Dabei obliegt dem Verwaltungsrat „als Kollektivorgan […] die Gestaltung der Geschäftspolitik und die damit verbundene Fassung grundlegender Beschlüsse für die Geschäftsführung".[126]

Das *kanadische* Gesellschaftsrecht setzte, ganz wie seine englischen Vorläufer, zunächst drei Direktoren im monistischen Leitungsorgan voraus.[127] Nachdem eine Regierungskommission in der Provinz Ontario allerdings die Streichung der gleichlautenden Provinzvorschrift (§ 296 II

---

[118] *Takahashi*, AG 2007, 476, 478; *Marutschke*, Einführung 2010, 254: „wohl die drastischste Reform in der Geschichte des japanischen Gesellschaftsrechts"; zu den Gründen ausf. *Menkhaus/Pawlowski* in: Süß/Wachter, Hdb GmbH 2011, 999 ff.

[119] *Takahashi*, AG 2007, 476, 479; *Behrendt*, Japan 2007, 56; *Marutschke*, Einführung 2010, 274; zum früheren Recht *Igarashi/Makino*, Einführung 1990, 150.

[120] *Meckel*, Corporate Governance 2010, 39.

[121] *Marutschke*, Einführung 2010, 256, 284 f.

[122] *Menkhaus/Pawlowski* in: Süß/Wachter, Hdb GmbH 2011, 1006.

[123] *Behrendt*, Japan 2007, 56.

[124] *Meckel*, Corporate Governance 2010, 42: „der Verwaltungsrat als Kollegialorgan bestätigte Entscheidungen dieses Ausschusses lediglich formal."; ebenso *Marutschke*, Einführung 2010, 277; *Kawamoto*, Japan 2004, 179 (Rn. 447); *Behrendt*, Japan 2007, 57.

[125] *Behrendt*, Japan 2007, 60 f.; *Menkhaus/Pawlowski* in: Süß/Wachter, Hdb GmbH 2011, 1005.

[126] *Meckel*, Corporate Governance 2010, 41; *Marutschke*, Einführung 2010, 274, 277: „Verwaltungsrat als Kollegialorgan"; *Menkhaus/Pawlowski* in: Süß/Wachter, Hdb GmbH 2011, 1007: „Form eines Kollegialorgans"; ebenso schon vor der Reform *Kawamoto*, Japan 2004, 253 (Rn. 656).

[127] Abschnitt 4 Joint Stock Companies Act 1850 (zit. nach *ECO*, Bill 14 / 1850); § 100 Dominion Companies Act 1918 (*Masten/Fraser*, Company Law 1929, 565); § 84 Dominion Companies Act 1934 (*Lighthall*, Companies Act 1935, 196).

R.S.O. 1960 ch. 71) empfohlen hatte, weil sie „keinen sinnvollen Zweck“ erfülle,[128] wurde die Mindestzahl in Ontario und im Bundesrecht aufgegeben. Das wurde auch damit gerechtfertigt, dass die Mindestzahl in großen Gesellschaften unnötig, weil ohnehin vorteilhaft, gewesen sei und in kleinen Gesellschaften zur Formalie verkommen sei, da der Hauptaktionär ohnehin dafür gesorgt habe, dass nur „Familienmitglieder, enge Freunde oder andere Ja-Sager zu Direktoren bestellt wurden“.[129] Bestehen blieb die Mindestzahl lediglich für Gesellschaften, die sich über den Kapitalmarkt finanzieren (§ 102 II CBCA).[130]

Das *österreichische* Gesellschaftsrecht lehnt sich eng an sein deutsches Vorbild an und setzt für die Geschäftsführer der GmbH (§ 18 öGmbHG) und den Vorstand der AG (§ 71 II öAktG) keine Mindestzahl voraus. „Einpersonenvorstände sind daher zulässig, aber selten.“[131] Der österreichische Corporate Governance Kodex empfiehlt dagegen die plurale Vorstandsbesetzung (Ziff. 16 S. 1 ÖCGK),[132] weshalb börsennotierte Aktiengesellschaften etwaige Abweichungen öffentlich erklären müssen (§ 243b I Nr. 3 UGB). Im Gegensatz zum AG-Vorstand zwingend plural muss dagegen das Leitungsorgan des österreichischen Vereins ausgestaltet werden; es muss seit 2002 „aus mindestens zwei Personen bestehen“ (§ 5 III 1 VerG). Auch die Aufsichtsorgane der drei Körperschaftsformen sind plural zu besetzen: Im Verein muss es (wenn statutarisch vorgesehen) „aus mindestens drei natürlichen Personen“ bestehen (§ 5 IV 1 VerG), in expliziter Anlehnung an die „gesellschaftsrechtlichen Grundsätze“[133] zur Besetzung des Aufsichtsrats in der großen GmbH mit drei (§§ 29 I, 30 S. 1 öGmbHG) und in der AG mit drei bis 20 Mitgliedern (§ 86 I öAktG).[134] Nach Sondergesetzen gelten in der AG mitunter höhere Mindestzahlen für den Aufsichtsrat,[135] und für den Vorstand im Finanzdienstleistungsbereich oft „eine Mindestan-

---

[128] 1.3.1 Report of the Select Committee on Company Law (unter Vorsitz von A. F. Lawrence), zit. nach *Lavine*, Corporations Act 1971, 2, 210.

[129] *Welling*, Corporate Law 1991, 309 f.

[130] Aktuelle Fassung unter http://laws-lois.justice.gc.ca/eng/acts/C-44/page-36.html#h-18; vgl. auch *Nockelmann* in: Süß/Wachter, Hdb GmbH 2011, 1045; *Welling*, Corporate Law 1991, 310 m.w.N. in Fn. 31 zu den entsprechenden Provinzvorschriften.

[131] *Kalss* in: Goette/Habersack, MK-AktG 2014, § 76 Rn. 135 m.w.N. in Fn. 520.

[132] Österreichischer Corporate Governance Kodex (ÖCGK) i.d.F. von Juli 2012 sowie frühere Fassungen unter www.wienerborse.at/corporate/kodex.htm.

[133] So öBReg, Begr. zu § 5 VerG-E, online in *Parlament*, NR: GP XXI RV 990, 2002, S. 26; genauer verweist *Krejci*, VerG 2002, § 5 Rn. 34 auf das „Recht der Kapitalgesellschaften (§ 86 Abs. 1 AktG; § 30 GmbHG).“

[134] Vgl. auch *Sattlegger* in: Ars Legis e.V., Kapitalgesellschaften 2007, 291; *Van Hulle/Gesell*, Corporate Law 2006, 63.

[135] Näher *Kalss* in: Doralt/Nowotny/Kalss, öAktG 2012, § 86 Rn. 12 ff.

zahl von zwei Vorstandsmitgliedern".[136] In jedem Fall sind mehrere Organmitglieder nur gemeinschaftlich geschäftsführungsbefugt,[137] mithin handeln Vorstand und Aufsichtsrat nach österreichischem Recht als Kollegialorgane.[138]

Das *polnische* Handelsgesellschaftengesetz macht keine Vorgaben zur Größe des Vorstands in der GmbH oder Aktiengesellschaft (Artt. 201, 368 § 2 KSH).[139] Sind allerdings mehrere Vorstandsmitglieder bestellt, so handeln sie grundsätzlich gemeinschaftlich (Artt. 205, 371 § 1 KSH).[140] Als Aufsichtsorgan sind in GmbH mit mindestens 25 Gesellschaftern und einem bestimmten Stammkapital sowie in den durch Privatisierung entstandenen ein Aufsichtsrat und/oder eine Prüfungskommission zu bilden (Art. 213 § 2 KSH),[141] die jeweils aus mindestens drei Mitgliedern bestehen müssen (Artt. 215, 217 KSH). Auch in Aktiengesellschaften muss der Aufsichtsrat aus mindestens drei, in Publikumsgesellschaften sogar fünf Mitgliedern bestehen (Art. 385 § 1 KSH).[142] Der Aufsichtsrat „erfüllt seine Pflichten kollegial".[143]

Im *russischen* Kapitalgesellschaftsrecht ist die auf 50 Gesellschafter beschränkte GmbH „die mit Abstand häufigste Rechtsform",[144] daneben gibt es aber auch die ebenfalls auf 50 Gesellschafter beschränkte geschlossene Aktiengesellschaft sowie die offene Aktiengesellschaft. In allen drei Rechtsformen liegt die Geschäftsführung grundsätzlich bei einem eingliedrigen Organ, das zwingend und ausschließlich einzelvertretungsberechtigt

---

[136] *Kalss* in: Kalss/Nowotny/Schauer, Hdb öGftsR 2008, Rn. 3/246 m.w.N. in Fn. 5; *Kalss* in: Goette/Habersack, MK-AktG 2014, § 76 Rn. 136.

[137] *Sattlegger* in: Ars Legis e.V., Kapitalgesellschaften 2007, 282 bzw. 290; *Van Hulle/Gesell*, Corporate Law 2006, 57 bzw. 62; zur GmbH *Beer* in: Süß/Wachter, Hdb GmbH 2011, 1302: „Mehrere Geschäftsführer haben die Geschäfte gemeinsam zu führen."

[138] *Kalss* in: Kalss/Nowotny/Schauer, Hdb öGftsR 2008, Rn. 3/339: „Nach dem Gesetz obliegt dem Gesamtvorstand als *Kollegium* die Geschäftsführung" und Rn. 3/532: „Der Aufsichtsrat ist ein Kollegialorgan"; ebenso *Kalss* in: Goette/Habersack, MK-AktG 2014, § 77 Rn. 70 und § 107 Rn. 195; ausf. auch *Resch*, GesRZ 2000, 2.

[139] Englische Fassung online unter www.roma.trade.gov.pl/it/download/file/f,6941; vgl. auch *Van Hulle/Gesell*, Corporate Law 2006, 285 bzw. 294; *Oplustil* in: Liebscher/Zoll, Einführung 2005, 410, 417 (Rn. 17) bzw. 430 (Rn. 53).

[140] Vgl. auch *Brockhuis/Schnell*, GftsR in Polen 2002, 64, 80; *Oplustil* in: Liebscher/Zoll, Einführung 2005, 410, 430 (Rn. 54): „Kollegialitätsprinzip, es sei denn, die Satzung der AG sieht etwas anderes vor (Art. 371)".

[141] *Brockhuis/Schnell*, GftsR in Polen 2002, 64, 110; *Bogen/Siekierzynski* in: Süß/Wachter, Hdb GmbH 2011, 1350.

[142] *Van Hulle/Gesell*, Corporate Law 2006, 295; *Oplustil* in: Liebscher/Zoll, Einführung 2005, 410, 431 (Rn. 56).

[143] *Oplustil* in: Liebscher/Zoll, Einführung 2005, 410, 431 (Rn. 58); wohl ebenso *Kuper*, Polen 2005, 57: „grundsätzlich kollegial besetzt".

[144] *Wedde*, OstEuR 2009, 154.

ist,[145] unbeschadet der gesetzlich zulässigen, aber fakultativen Einrichtung eines zusätzlichen kollegialen „Exekutivorgans".[146] Sowohl in der GmbH mit mehr als 15 Gesellschaftern als auch in der AG „ist als internes Prüfungsorgan eine Revisionskommission zu bilden, die Kollegial- oder Einzelorgan sein kann."[147] Als viertes Gesellschaftsorgan schließlich kennt das russische Recht den „mindestens siebenköpfigen"[148] Direktoren- oder Aufsichtsrat, der in GmbH und Aktiengesellschaften mit weniger als 50 Gesellschaftern freiwillig, im übrigen aber zwingend einzurichten ist.[149]

Das *schwedische* Aktiengesetz als Mutter des finnischen wurde wie jenes 2006 novelliert.[150] Es kennt keine GmbH und keinen Aufsichtsrat, sondern nur die monistisch organisierte Aktiengesellschaft in einer Publikums- und einer privaten Variante.[151] Wie das finnische Aktienrecht erforderte auch das schwedische früher grundsätzlich ein dreigliedriges Verwaltungsorgan und sah lediglich eine Ausnahme für kleine Gesellschaften vor (ABL 8:1 a.F.).[152] Inzwischen ist das Regel-Ausnahme-Verhältnis umgekehrt: Grundsätzlich genügt ein Mitglied mit Stellvertreter (ABL 8:1), während nach einer Sondervorschrift für Publikumsgesellschaften mindestens drei Mitglieder zu bestellen sind (ABL 8:46).[153] In Publikumsgesellschaften muss das Leitungsorgan einen geschäftsführenden Direktor bestellen, dessen Stellung „vergleichbar [mit] derjenigen eines deutschen Vorstands" ist.[154]

Im *schweizerischen* Gesellschaftsrecht, dessen Reform 2008 „die erste Totalrevision des GmbH-Rechts seit [Einführung der GmbH] 1936" be-

---

[145] *Pashchenko* in: Nußberger, Einführung 2010, 216, 226.

[146] *Radjuk*, RIW 2009, 592, 595 f. und *Görlitz* in: Süß/Wachter, Hdb GmbH 2011, 1444 (für die GmbH); *Schmitt/Vogt*, RIW 2002, 762, 763 (für die AG); *Pashchenko* in: Nußberger, Einführung 2010, 216, 226, 231.

[147] Für die GmbH *Pashchenko* in: Nußberger, Einführung 2010, 216, 226 und *Wedde*, OstEuR 2009, 154, 162; für die AG *Schmitt/Vogt*, RIW 2002, 762, 763.

[148] *Lüdemann*, Russland 2001, 147.

[149] Für die GmbH *Radjuk*, RIW 2009, 592, 596 und *Wedde*, OstEuR 2009, 154, 162; für die AG *Schmitt/Vogt*, RIW 2002, 762, 763.

[150] *Foerster/Kastner* in: Süß/Wachter, Hdb GmbH 2011, 1454; *Nicolaysen*, RIW 2005, 884, passim (auf 885 mit dem Hinweis auf die zunehmende Divergenz der früher gleichlaufenden skandinavischen Rechtsordnungen).

[151] *Sjöstedt* in: Ars Legis e.V., Kapitalgesellschaften 2007, 297 f.; *Nicolaysen*, RIW 2005, 884, 885 f.

[152] *Ring/Olsen-Ring*, Einführung 1999, 182; deutsche Fassung abgedr. bei *Carsten/Korkisch*, Schwedisches AktR 1986, 108.

[153] Englische Fassung abgedr. bei *Skog/Fäger*, Companies Act 2007, 197, 207; vgl. auch *Nicolaysen*, RIW 2005, 884, 886; *Van Hulle/Gesell*, Corporate Law 2006, 350, 357; *Foerster/Kastner* in: Süß/Wachter, Hdb GmbH 2011, 1478.

[154] *Nicolaysen*, RIW 2005, 884, 886; vgl. auch *Foerster/Kastner* in: Süß/Wachter, Hdb GmbH 2011, 1478.

wirkte,[155] gilt für die GmbH grundsätzlich Selbstorganschaft (Art. 809 I 1 OR) und auch für den Vereinsvorstand gibt es keine Organisationsvorgaben (vgl. Art. 69 ZGB). Für den Verwaltungsrat der Aktiengesellschaft – „in der Schweiz die bei weitem bedeutendste und häufigste Rechtsform für irgendeine wirtschaftliche Tätigkeit“[156] – wurden zwar in der Aktienrechtsrevision 1991 Größenvorgaben diskutiert, letztlich aber verworfen,[157] daher genügt heute ein Einmann-Verwaltungsrat (Art. 707 I OR).[158] Gleichwohl legt dessen gesetzliche Aufgabenzuweisung (Art. 716a I OR) nahe, dass „in einer arbeitsteilig organisierten, unternehmerisch tätigen Gesellschaft praktisch ein *mehrgliedriger Verwaltungsrat* notwendig“ ist,[159] weshalb der Verwaltungsrat nach herrschender Meinung „aus mindestens drei Personen bestehen, d.h. ein Kollegium bilden“ muss;[160] für Bankaktiengesellschaften ist das sogar spezialgesetzlich angeordnet.[161] Jedenfalls steht die Geschäftsführung grundsätzlich „allen Mitgliedern des Verwaltungsrates gesamthaft zu“ (Art. 716b III OR), so „dass im Zweifel das Kollegialprinzip gelten soll“.[162]

Das *spanische* „Kapitalgesellschaftsgesetz“ von 2010 ermöglicht GmbH und AG die Auswahl unter vier Strukturvarianten, darunter auch die Führung durch einen mindestens dreiköpfigen „Verwaltungsrat“ (Artt. 210 I, 242 I LSC).[163] Der Verwaltungsrat, der in der AG „dem Normalfall entspricht“, weil die anderen Gestaltungen „unüblich“ oder „nur vereinzelt üblich“ sind,[164] war schon in früheren Gesetzesfassungen vorgesehen.[165] Zwingend einzurichten ist er aber bis heute nur nach Sondergesetzen, insbesondere seit 2002 in börsennotierten Aktiengesellschaften,[166] während er

---

[155] *Schindler/Töndury* in: Süß/Wachter, Hdb GmbH 2011, 1498.

[156] *Forstmoser*, ZGR 2003, 688, 693 (ebenso schon 691).

[157] *Roth Pellanda*, Organisation 2007, 205 (Rn. 416) m.w.N. in Fn. 1360.

[158] Vgl. auch *Bühler/Etter* in: Ars Legis e.V., Kapitalgesellschaften 2007, 315 bzw. 313.

[159] *Böckli*, Schweizer AktR 2009, 1558 (zuvor schon 1550); ebenso *Roth Pellanda*, Organisation 2007, 207 (Rn. 420).

[160] *Böckli*, Schweizer AktR 2009, 1732; zust. *Bauen/Venturi*, Verwaltungsrat 2007, 190 (Rn. 505) bei Fn. 200; *Roth Pellanda*, Organisation 2007, 206 (Rn. 416).

[161] Art. 8 I BankV, zit. nach *Roth Pellanda*, Organisation 2007, 206 (Rn. 416).

[162] *Forstmoser*, ZGR 2003, 688, 694, weiter 695: Praktisch gelte das „nur in kleinen und kleinsten Verhältnissen“.

[163] *Dorfmeister* in: Ars Legis e.V., Kapitalgesellschaften 2007, 344; *Van Hulle/Gesell*, Corporate Law 2006, 337 f., 344; *Grechenig*, Verwaltungssystem 2005, 23; *Janka*, Spanien 2011, 66; dt. Fassung des *Ley de Sociedades de Capital* bei *Sohst*, Spanisches GftsR 2012, 52, 58.

[164] *Cremades*, GftsR in Spanien 2001, 44; *Janka*, Spanien 2011, 66: der Verwaltungsrat war „in der Praxis bereits der Regelfall“.

[165] Art. 73 LSA 1951, abgedr. bei *von Hoffmann/von Waldheim*, Spanisches AktR 1975, 112 („Modell kollegialer Verwaltung“, 28); später Art. 136 LSA 1989, abgedr. bei *Sohst*, Aktien- und GmbH-Gesetz 2002, 34.

in der GmbH überhaupt nur „selten“ anzutreffen ist.[167] Der Verwaltungsrat „entscheidet […] als Kollegialorgan durch mehrheitlichen Beschluss“,[168] überweist aber viele Befugnisse entweder an mehrgliedrige Ausschüsse oder an „abgeleitete Verwalter“ aus der Mitte des Verwaltungsrates.[169] Außer durch einen Verwaltungsrat kann die spanische AG auch durch mehrere Verwalter geführt werden, die jeweils entweder Gesamt- oder Einzelgeschäftsführungs- und -vertretungsbefugnis haben.[170] Die Einrichtung eines Aufsichtsrats ist nach spanischem Recht nicht möglich.[171]

Die *südkoreanische* Aktiengesellschaft als „übliche Gesellschaftsform für die großen und mittleren Unternehmen“[172] hatte nach dem bis 1962 in Südkorea geltenden japanischen Handelsgesetzbuch zwar „Vorsteher und Aufseher, aber keinen Vorstand und keinen Aufsichtsrat als Kollegialorgan.“[173] Erst mit dem koreanischen Handelsgesetzbuch von 1962 wurde nach US-amerikanischem Vorbild ein aus aus mindestens drei „Direktoren“ bestehender „Board“ eingeführt (§ 383 I 1 kHGB).[174] „Er ist grundsätzlich ein Kollegialorgan“,[175] bestellt aber für die Ausführung seiner Beschlüsse einen „Repräsentativ-Direktor“, der als selbstständiges Geschäftsführungsorgan neben dem Board fungiert.[176] Um die Geschäftsführung durch Board und Repräsentativ-Direktor zu überwachen, wurde der frühere interne Rechnungsprüfer 1984 zur „Aufsichtsperson“ aufgewertet (§§ 409 ff. kHGB),[177] deren jede Aktiengesellschaft eine oder mehrere haben muss.[178]

---

[166] *Grechenig*, Verwaltungssystem 2005, 23 mit Fn. 191, 24 bei Fn. 202; *Janka*, Spanien 2011, 66 f. mit Fn. 208.

[167] *Grechenig*, Verwaltungssystem 2005, 196.

[168] *Janka*, Spanien 2011, 68; *Cremades*, GftsR in Spanien 2001, 46: „ein Kollegialorgan“; *Grechenig*, Verwaltungssystem 2005, 35: „kollegial organisierte Gesellschaftsverwaltung“.

[169] *Adomeit/Frühbeck*, Einführung 2001, 113; *Grechenig*, Verwaltungssystem 2005, 36; *Janka*, Spanien 2011, 68 f.

[170] *Janka*, Spanien 2011, 66; *Löber/Lozano/Steinmetz* in: Süß/Wachter, Hdb GmbH 2011, 1655, 1696; *Grechenig*, Verwaltungssystem 2005, 23.

[171] *Grechenig*, Verwaltungssystem 2005, 23 Fn. 188, 196 m.w.N. in Fn. 1885.

[172] *Lee* in: Dauner-Lieb u.a., FS Konzen 2006, 491, auf 492 mit einer Statistik von 2004: 85,5 % der Handelsgesellschaften waren Aktiengesellschaften, nur 3,7 % GmbH.

[173] *Yi*, Strukturreform 2004, 188 mit ausf. Erläuterung in Fn. 1110.

[174] *Yi*, Strukturreform 2004, 190 f., 210; *Schleicher*, AG koreanischen Rechts 2007, 56; deutsche Fassung in *Son*, Koranisches AktR 1989, 92.

[175] *Welling*, Corporate Law 1991, 56; *Yi*, Strukturreform 2004, 210: „Der Board ist ein Kollegialorgan, das aus den Direktoren besteht.“

[176] *Yi*, Strukturreform 2004, 210 f.; *Lee*, Geschäftsführung 1991, 41 sieht das darin begründet, dass der Board als „Kollegialorgan für die Wahrnehmung der täglichen Geschäftsführungs- und Vertretungsaufgaben zu schwerfällig“ sei.

[177] *Yi*, Strukturreform 2004, 194; deutsche Fassung der einschlägigen §§ 409 ff. kHGB bei *Son*, Koranisches AktR 1989, 97 (dort aber als „Aufsichtsrat“ übersetzt).

[178] *Yi*, Strukturreform 2004, 224: „Es muss […] mindestens eine Aufsichtsperson vorhanden sein, obwohl das KHGB eine Mindestzahl nicht ausdrücklich vorschreibt. […] Das

„Dieses besondere Verwaltungssystem ist auch dadurch gekennzeichnet, […] dass mehrere Aufsichtspersonen kein Kollegium bilden. Sie sind also keine Mitglieder eines Kollegialorgans, sondern selbstständige Organe, so wie auch Vorsteher und Aufseher“ nach dem bis 1962 geltenden japanischen HGB.[179] Seit 1999 – nach der Finanz- und Währungskrise[180] – haben kleine Aktiengesellschaften ein Wahlrecht und große Aktiengesellschaften (Vermögen umgerechnet mindestens 2 Mio. Euro) die Pflicht, anstelle der Aufsichtsperson einen „Prüfungsausschuss“ einzurichten, der aus mindestens drei Personen bestehen muss (§ 415-2 II 1 kHGB).[181] Andere Ausschüsse sind fakultativ, müssen aber je mindestens zwei Mitglieder umfassen (§ 393-2 III kHGB). „Der Prüfungsausschuss ist ein Kollegium“, aber kein eigenständiges Organ der Gesellschaft, sondern lediglich ein „Ausschuss innerhalb des Vorstands“.[182] Daher bildet die Aufsichtsperson nach wie vor das „eigentliche und Haupt-Überwachungsorgan der koreanischen AG“.[183] Ebenfalls im Zuge der Reform 1998 wurde für kleine Aktiengesellschaften (Grundkapital umgerechnet höchstens 350.000 Euro) die Mindestgröße des Board aufgegeben, seither genügen in diesen Fällen also ein oder zwei Direktoren (§ 383 I Hs. 2 kHGB).[184]

Die *US-amerikanische* Aktiengesellschaft, die durch ein monistisches Leitungsorgan (*board*) geführt wird, war „seit Menschengedenken […] verpflichtet, drei Direktoren zu haben.“[185] Die entsprechende Gesetzgebungskompetenz liegt bei den Bundesstaaten, und bis 1961 hatten alle außer Arizona, Iowa, Rhode Island und South Dakota eine entsprechende Vorschrift,[186] ebenso wie das 1950 zur Rechtsvereinheitlichung von der

---

KHGB sieht auch keine Höchstzahl der Aufsichtspersonen vor.“; *Lee*, Geschäftsführung 1991, 47: „In der Praxis haben Aktiengesellschaften zumeist nur ein Aufsichtsratsmitglied.“

[179] *Yi*, Strukturreform 2004, 196, vgl. auch 218 (genauso 224 bzw. 328 ff.): „Wenn mehrere Aufsichtspersonen bestellt sind, bilden sie kein Kollegialorgan. Vielmehr ist jede Aufsichtsperson für sich ein Organ. Die Einrichtung des Kollegialorgans der Aufsichtspersonen ist aber erlaubt.“; ebenso *Lee*, Geschäftsführung 1991, 47.

[180] Näher *Lee* in: Dauner-Lieb u.a., FS Konzen 2006, 491, 493.

[181] *Yi*, Strukturreform 2004, 217, 219; engl. Fassung online in *MOLEG*, Commercial Act 2009, 111; dagegen spricht *Lee* in: Dauner-Lieb u.a., FS Konzen 2006, 491, 497 von „mindestens vier“ Mitgliedern und sieht nur „die zum Börsenhandel zugelassenen Aktiengesellschaften“ zur Einrichtung des Ausschusses verpflichtet an.

[182] *Yi*, Strukturreform 2004, 219 f. (dort auf 220 erstes Zitat); *Schleicher*, AG koreanischen Rechts 2007, 57 (dort zweites Zitat).

[183] *Yi*, Strukturreform 2004, 223 (Kapitelüberschrift zu § 9).

[184] *Lee* in: Dauner-Lieb u.a., FS Konzen 2006, 491, 495.

[185] *Spoerri*, Bus Lawyer 1963, 305, 307; ebenso *Rudolph*, Bus Lawyer 1965, 781: “a long standing and almost universal corporate norm.”; *ABA*, MBCA Annotated 2011, 8-34: “The historical tradition, going back to Blackstone or earlier”; *Pollman*, UC Davis Bus L J 2009, 137, 140: “historic rule and prevailing norm”, aber “not generally required”.

[186] *Spoerri*, Bus Lawyer 1963, 305, 307.

US-Anwaltskammer (ABA) entworfene Mustergesetz (§ 8.03 (a) MBCA).[187] Nachdem die formale Mindestgröße in der Praxis aber oft nur zum Schein eingehalten wurde,[188] schaffte 1961 zuerst Wyoming, kurz darauf Delaware,[189] die Mindestzahl ab. 1969 folgte das Mustergesetz diesem Vorbild, „in erster Linie, um eigentümergeführten oder *joint venture* Aktiengesellschaften mit einem oder zwei Aktionären zu ermöglichen, den Vorstand ausschließlich mit Personen zu besetzen, die an der Gesellschaft beteiligt sind."[190] Heute schreiben nur noch sechs,[191] inzwischen wohl sogar nur noch drei,[192] Bundesstaaten ein plural besetztes Leitungsorgan vor (§ 212 (a) 2 Hs. 1 Cal. Corp. Code; Ch. 156D § 8.03 (a) Mass. Gen. Laws; § 16-10a-803 (1) (a) Utah Code).[193] Doch sogar „alle diese Jurisdiktionen haben auch eine Vorschrift, nach der unter bestimmten Umständen weniger Direktoren bestellt werden können."[194] Zwei weitere Bundesstaaten sehen als Regelfall drei Direktoren vor, wenn Satzung und Geschäftsordnung keine andere Zahl vorgeben (§ 10.06.543 (a) 3 Alaska Stat.; Ch. 15 § 1723 (a) 3 Pa. Cons. Stat.), die übrigen 43 Staaten lassen das Ein-Mann-*board* unbeschränkt zu.[195] Anders stellt sich die Situation allerdings im Recht der gemeinnützigen Unternehmen dar. Dort ist die mindestens dreigliedrige Besetzung des Leitungsorgans durch zahlreiche Bundesstaaten sowie im neuen Mustergesetz der Anwaltskammer vorgeschrieben (§ 8.03 (a) MNCA).[196] Für börsennotierte Unternehmen gilt letztlich dieselbe Mindestgröße, da die Börsenzulassungsregeln einen mindestens dreiköpfigen Prü-

---

[187] *ABA*, MBCA Annotated 2011, 8-32.

[188] *Spoerri*, Bus Lawyer 1963, 305, 308: "notorious practices of employing dummy or accommodation incorporators and directors, and of recording and certifying minutes of meetings that are not held."; krit. *Garrett*, Bus Lawyer 1964, 535, 537.

[189] *Rudolph*, Bus Lawyer 1965, 781 f.

[190] *ABA*, MBCA Annotated 2011, 8-34, ebenso 8-32, dort aber mit dem Zusatz "or for corporations with more than two shareholders where in fact the full power of management is vested in only one or two persons."

[191] *ABA*, MBCA Annotated 2011, 8-37 nennt "California, Massachusetts, Missouri, Ohio, Utah, and Vermont".

[192] Von den Staaten in der vorigen Fn. fallen laut Online-Fassung ihrer Gesetze (Zugriff am 10.10.2012) Missouri, Ohio und Vermont weg: Nach § 351.315 (1) 1 Mo. Stat. (www.moga.mo.gov/statutes/C300-399/3510000315.HTM) sowie Ch. 11A § 8.03 (a) 1 Vt. Stat. (www.leg.state.vt.us/statutes/fullchapter.cfm?Title=11a&Chapter=008) genügen "one or more individuals", nach § 1701.56 (A) (1) Ohio Rev. Code "not [.] less than one" (codes.ohio.gov/orc/1701.56).

[193] Für Puerto Rico vgl. *ABA*, MBCA Annotated 2011, 8-37 ("requires the board to be made up of two or more members."), 8-36 m.Verw. auf 14 P.R. Laws Ann. §§ 2721, 2915.

[194] *ABA*, MBCA Annotated 2011, 8-37.

[195] Liste in *ABA*, MBCA Annotated 2011, 8-37; *Bungert*, GftsR in den USA 2003, 25.

[196] *ABA*, MNCA 2009, zit. nach Entwurfsfassung (http://ali.state.al.us/docs/Nonprofit-Corp-Act-8-2008.pdf).

fungsausschuss (*audit committee*) vorschreiben.[197] In einer solchen pluralen Besetzung „stellt das *board* ein Kollegialorgan dar“[198] und handelt „grundsätzlich als Organ, nicht durch einzelne Mitglieder.“[199]

Das Gesellschaftsrecht im *Vereinigten Königreich* gehört zu den traditionsreichsten. Schon das erste Aktiengesetz von 1844 erforderte die Bestellung von mindestens drei Direktoren.[200] Später genügte die zweigliedrige Unternehmensleitung – durch mindestens zwei Direktoren in öffentlichen Gesellschaften[201] bzw. einen Direktor und einen Gesellschaftssekretär in geschlossenen Gesellschaften.[202] Mit der letzten großen Gesellschaftsrechtsreform 2006 wurde die Organpluralität noch weiter zurückgenommen und geschlossenen Gesellschaften die eingliedrige Unternehmensführung ermöglicht.[203] Zugleich wird davon allerdings abgeraten.[204] Öffentliche Gesellschaften müssen nach wie vor mindestens zweigliedrig geleitet werden (§ 154 II CA 2006).[205] Für börsennotierte Gesellschaften empfiehlt der englische Corporate Governance Kodex sogar die Einrichtung eines Prüfungs- sowie eines Vergütungsausschusses von jeweils mindestens zwei, in den Unternehmen des FTSE 350 sogar mindestens drei Mitgliedern (C.3.1, D.2.1 CGC).[206] Nach den Mustersatzungen, die in englischen Kapitalgesellschaften subsidiär zu den jeweils vereinbarten Statuten gelten (§ 20 I CA 2006), erfolgt die Geschäftsführung sowohl in der geschlossenen als auch in der offenen Gesellschaft durch kollektive Entscheidung.[207] „Auch in England wird das *board* als Kollegialorgan begriffen“ und ob-

[197] *NYSE*, Listed Company Manual 2012, 303A.06, 303A.07 (a) 1.

[198] *Beckert*, Personalisierte Leitung 2009, 101 Fn. 122 m.w.N. in Fn. 614; *Gerber* in: Süß/Wachter, Hdb GmbH 2011, 2015.

[199] *Merkt/Göthel*, US GftsR 2006, 319 (Rn. 586); *Garrett*, Bus Lawyer 1964, 535: “Most of the statutory corporation law is framed around ‘togetherness’ in corporate management”.

[200] § 110 Joint Stock Companies Act 1844 (7 & 8 Vict.), zit. nach *Palmer* in: Ziegel, Studies I 1967, 365, 366 Fn. 7.

[201] Vgl. § 139 I CA 1929, § 176 CA 1948 und § 282 CA 1985.

[202] Vgl. zuletzt § 283 I, II CA 1985: “Every Company shall have a secretary. A sole director shall not also be secretary.”

[203] § 270 I CA 2006: “A private company is not required to have a secretary.”; *Elwell* in: Ars Legis e.V., Kapitalgesellschaften 2007, 128; *Van Hulle/Gesell*, Corporate Law 2006, 361.

[204] *Koh Soon Kwang*, Company Law 2010, 118 (Rn. 5.66): “Although it is sufficient just to have 1 director for private companies, it is not advisable to do so.”

[205] Vgl. auch *Elwell* in: Ars Legis e.V., Kapitalgesellschaften 2007, 128; *Köritz*, Konvergenz 2010, 393.

[206] *FRC*, UK CGC 2012, 18 bzw. 22; zum FTSE 350 oben § 3 Fn. 246.

[207] Ziff. 7 I Model Articles (pcls): “The general rule about decision-making by directors is that any decision of the directors must be either a majority decision at a meeting or a [unanimous] decision”, ähnl. Ziff. 7 Model Articles (plc), beide online in *CH*, Model Articles 2008, 5 bzw. 35.

wohl eine anderweitige Satzungsgestaltung möglich ist, scheint sie „in der Praxis bisher nicht vorgekommen“ zu sein.[208]

Nach diesem Streifzug durch 19 Rechtsordnungen betrachte ich zuletzt mit der *Societas Europaea (SE)* eine transnationale Rechtsform, die 2004 auf europäischer Ebene eingeführt wurde. Die SE darf grundsätzlich zwischen der monistischen und dualistischen Struktur wählen und unterliegt neben dem europäischen Regelungsrahmen auch dem Organisationsrecht ihres Sitzstaates, auf das ich hier nicht eingehe.[209]

Zur Organpluralität macht das europäische Recht keine Vorgaben – in der *dualistischen* SE weder für das Leitungs- (Art. 39 IV SEVO) noch für das Aufsichtsorgan (Art. 40 III 1 SEVO). Lediglich in Vorentwürfen war das Aufsichtsorgan auf eine durch drei teilbare Mitgliederzahl beschränkt und musste sogar mindestens zwölf Mitglieder umfassen „wenn die S.E. Betriebsstätten in mehreren Mitgliedstaaten hat.“[210] Auch für die *monistische* SE hatten Vorentwürfe ein mindestens dreigliedriges Verwaltungsorgan vorgesehen (zuerst ausnahmslos,[211] danach immerhin im Grundsatz[212]), doch heute ist die Größe des Verwaltungsorgans unbeschränkt (Art. 43 II 1 SEVO) mit Ausnahme des mitbestimmten Verwaltungsrats, der mindestens drei Mitglieder haben muss (Art. 43 II UA 2 SEVO).[213]

Auch die Organkollektivität war in Vorentwürfen noch ausdrücklich geregelt[214] und ergibt sich heute nur noch mittelbar aus den Artt. 43 I 1, 50 I SEVO für den Verwaltungsrat der monistischen und aus Artt. 39 I 1, 50 I,

---

[208] *Beckert*, Personalisierte Leitung 2009, 101 Fn. 125 m.w.N. in Fn. 645; auch nach *Köritz*, Konvergenz 2010, 399 f. „liegt dem englischen Recht das Kollegialprinzip zugrunde“, von dem „durch die *articles of association* Abstand genommen werden kann“.

[209] In Deutschland sind die Leitungsorgane der dualistischen SE analog zu Vorstand und Aufsichtsrat der Aktiengesellschaft ausgestaltet (§§ 16 S. 1, 17 I SEAG), der monistische Verwaltungsrat als Hybrid aus Vorstand und Aufsichtsrat (§ 23 SEAG kombiniert § 16 S. 1 und § 17 I 1, I 4, II SEAG).

[210] Art. 74 I VOV 1970, vgl. *Europäische Kommission*, ABl. C 124 v. 10.10. 1970, 17; ebenso Art. 74 III SE-VOV 1975, vgl. *Europäische Kommission*, Bull. 4 / 1975, 44.

[211] Art. 66 I 2 SE-VOV 1989 („Das Verwaltungsorgan […] muss aus mindestens drei Mitgliedern bestehen.“), online in *Europäische Kommission*, ABl. C 176 v. 8.7. 1991, 1, 37.

[212] Art. 66 Ia 1 SE-VOV 1991 („das Verwaltungsorgan besteht aus mindestens drei Mitgliedern […] kann jedoch aus nur zwei Mitgliedern oder aus nur einem einzigen Mitglied bestehen, wenn die Vertretung der Arbeitnehmer in der SE nicht […] geregelt ist.“), online in *Europäische Kommission*, ABl. C 176 v. 8.7. 1991, 1, 37.

[213] *Merkt*, ZGR 2003, 650, 654 f.; dagegen *Lutter/Kollmorgen/Feldhaus*, BB 2005, 2473, 2474: „Verwaltungsrat mit im Regelfall drei Mitgliedern.“; vgl. auch *Beckert*, Personalisierte Leitung 2009, 137 zur Frage, inwieweit die gesetzliche Mitbestimmung eine kollegiale Organbesetzung erfordert.

[214] Art. 64 II 1 SE-VOV 1970: „Besteht der Vorstand aus mehreren Mitgliedern, so bilden diese ein Kollegium.“ (*Europäische Kommission*, ABl. C 124 v. 10.10. 1970, 15); ebenso später Art. 64 II SE-VOV 1975 (*Europäische Kommission*, Bull. 4 / 1975, 39).

II SEVO für den Vorstand der dualistischen SE.[215] Dementsprechend „gilt also auch für den Vorstand in der dualistischen SE das Kollegialprinzip, auch wenn der Verordnungstext dies weniger deutlich als das deutsche Aktiengesetz zum Ausdruck bringt."[216]

4. Zwischenergebnis: Was und wie verbreitet ist das Kollegialprinzip?

Als „Kollegialprinzip" können und werden zwei verschiedene Grundsätze bezeichnet: der Grundsatz der Organpluralität (Organe werden mit mehreren Entscheidungsträgern besetzt.) und der Grundsatz der Organkollektivität (Mehrere Entscheidungsträger gehen durch grundsätzlich gemeinsame und gleichberechtigte Beschlusstätigkeit jeweils gleiche Verpflichtungen ein.) Dieses Verständnis teilt grundsätzlich auch das Verwaltungs- und Gerichtsorganisationsrecht, wenngleich es erst dreigliedrige Gremien als „kollegial" bezeichnet.

Das in Deutschland ausgeprägte „organisationsrechtliche Vorverständnis", wonach Organpluralität und Organkollektivität zwingend gleichlaufen,[217] ist nicht universell. Das belegt der Rechtsvergleich, etwa mit Japan, Spanien und Südkorea, wo tatsächlich plural besetzte Organe gleichwohl nicht kollektiv handeln müssen,[218] oder mit Österreich, wo andersherum kollektives Organhandeln sogar unter nicht-pluralen Organen stattfinden soll.[219] Organpluralität und Organkollektivität als Dimensionen des „Kollegialprinzips" müssen daher gedanklich getrennt werden, um begriffliche Verwirrung zu vermeiden.

Im Übrigen lassen sich aus der Betrachtung von zwanzig Rechtsordnungen die folgenden Aussagen zur Verbreitung des Kollegialprinzips ableiten:

---

[215] *Beckert*, Personalisierte Leitung 2009, 115 bzw. 134 (m.w.N. auf 133 bei Fn. 709); ebenso *Reichert/Brandes* in: Goette/Habersack, MK-AktG 2012, Art. 43 SEVO Rn. 5 und wohl auch *Lutter/Kollmorgen/Feldhaus*, BB 2005, 2473, 2479; a.A. *Thamm*, NZG 2008, 132, 133: „Eine pauschale Charakterisierung der geschäftsführenden Direktoren als Kollegium Gleichberechtigter ist voreilig."

[216] *Beckert*, Personalisierte Leitung 2009, 104; ebenso *Teichmann* in: Lutter/Hommelhoff, SE-Kommentar 2008, Art. 50 SEVO Rn. 7: „Folglich geht die Verordnung davon aus, dass die Organkompetenzen von allen Mitgliedern im Grundsatz *gleichberechtigt* und mit gleicher Verantwortlichkeit wahrgenommen werden."

[217] *Hoffmann-Becking*, NZG 2003, 745, 747.

[218] Entsprechend trennt *Yi*, Strukturreform 2004, 293 beide Aspekte durch ein Semikolon: „Unter dem Kollegialprinzip sind die Mitglieder eines Organs gleichberechtigt und treffen Entscheidungen nach dem Einstimmigkeits- oder Mehrheitsgrundsatz; das Organ ist eine Pluralinstanz."

[219] So *Krejci*, VerG 2002, § 5 Rn. 22: „Dem Vier-Augen-Prinzip entspricht auch, wenn es einen ‚Präsidenten' und einen eingliedrigen ‚Vorstand' gäbe. Die Verflochtenheit der Geschäftsführungs- und Vertretungsaufgaben führt [.] zu einer so dichten Zusammenarbeit der beiden Organe, dass das Ziel des Vier-Augen-Prinzips auch in diesem Fall erreicht wird."

1. Börsennotierte Unternehmen sind durchweg verpflichtet oder durch informelle Regelwerke angehalten, ihre Führungsorgane plural zu besetzen. Kleine und geschlossene Unternehmen hingegen sind nirgends zur pluralen Besetzung ihrer Organe verpflichtet.
2. In dualistischen Systemen der Unternehmensführung muss das Aufsichtsorgan durchweg plural besetzt werden, das Leitungsorgan dagegen nur in etwa der Hälfte der Rechtsordnungen.
3. Im monistischen System der Unternehmensführung ist für das Verwaltungsorgan ganz überwiegend die plurale Besetzung vorgeschrieben, mit Ausnahme der Schweiz und USA.
4. Über gemeinnützige Körperschaften ist keine allgemeine Aussage möglich: Das österreichische Vereinsrecht und US-amerikanische Recht der gemeinnützigen Unternehmen schreiben Organpluralität vor, das deutsche und schweizerische Vereinsrecht verzichten auf Vorgaben.
5. Regelmäßig müssen plural besetzte Organe kollektiv handeln, aber einzelne Organgestaltungen (Verwalter in Spanien, Aufseher in Südkorea) durchbrechen diesen Gleichlauf.
6. Die jüngste Reformwelle um 2006, in der ein Drittel der hier betrachteten Gesellschaftsrechte novelliert wurden, hat die rechtliche Organpluralität durchweg reduziert.

Zusammenfassend lässt sich also feststellen, dass zwar in keiner Rechtsordnung die Leitungsorgane von Kapitalgesellschaften ausnahmslos plural besetzt und kollektiv ausgestaltet sein müssen, dass es aber auch keine Rechtsordnung gibt, die das Kollegialprinzip gar nicht kennt.

II. Zweckerwägungen im systematischen und internationalen Vergleich

Nachdem die große Verbreitung des Kollegialprinzips in seinen zwei Ausprägungen festgestellt ist, sind nun die dafür ins Feld geführten Zweckerwägungen zu ermitteln. Die Auslegung nach Sinn und Zweck ist sowohl einer der wichtigsten Auslegungsschritte als auch die wichtigste Schnittstelle zwischen Rechtsdogmatik und empirischer Forschung (§ 1 F.I.).

Im Folgenden werden Zweckerwägungen aus der deutsch- und englischsprachigen Literatur berücksichtigt. Dabei ist in Deutschland auch das Recht der Verwaltungs- und Justizorganisation einzubeziehen, da bisher viel dafür spricht, dass „das Kollegialprinzip als Rechtsprinzip eines allgemeinen Organisationsrechts […] in allen Rechtsgebieten im Privatrecht wie im öffentlichen Recht“[220] ähnliche Zwecke verfolgen könnte.

---

[220] *Hoffmann-Becking*, NZG 2003, 745, 746; *Hoffmann-Becking*, ZGR 1998, 497, 507.

1. Deutsches Aktienrecht

Erste Anhaltspunkte für Sinn und Zweck des Kollegialprinzips bietet die Entwicklungsgeschichte des deutschen Aktienrechts. Zur *Organpluralität* finden sich in den Materialien zum AktG 1965 drei Zweckerwägungen: dass erstens „in größeren Gesellschaften […] die Arbeitslast des Vorstands meist die Kraft einer Person übersteigt"; zweitens sei „die Verantwortung für die geschäftlichen Maßnahmen häufig so groß, dass sie auf mehrere Personen verteilt werden sollte"; drittens schließlich „hat ein mehrköpfiger Vorstand den Vorzug, dass die Mitglieder sich wegen ihrer gesamtschuldnerischen Haftung wechselseitig überwachen und dadurch die Gefahr von Missbräuchen verringert wird."[221] Letzteres war bereits in der geschichtlichen Entwicklung des Kollegialprinzips eine maßgebliche Erwägung: Der absolutistische Herrscher nutzte Kollegialbehörden nicht zuletzt dazu,

> „sich […] der steigenden Übermacht des Fachwissens zu erwehren und ihm gegenüber in seiner Herrenstellung zu behaupten [.]. Er hält einen Fachmann durch andere im Schach und sucht sich […] die Gewißheit zu verschaffen, daß ihm nicht willkürliche Entscheidungen souffliert werden."[222]

Dementsprechend solle auch das Gesellschaftsrecht „nicht durch Rechtsmacht, sondern Psychologie" eine gegenseitige Kontrolle der Organmitglieder bewirken.[223] Zu den „klar erkennbaren Vorteilen des Kollegialorgans" zählten vor allem „Fehlerausgleich, Ausgewogenheit und Selbstprüfung der gefällten Entscheidungen",[224] daher ergänze das Kollegialprinzip

> „die vertikale (Fremd-)Kontrolle des Aufsichtsrates um eine horizontale (Selbst-)Kontrolle des Vorstands. Diese zusätzliche Kontrollebene ist um so wichtiger, je weniger der Aufsichtsrat in der Lage ist, seine Überwachungsaufgaben wirkungsvoll auszuüben."[225]

---

[221] Begr. § 76 AktG-RegE 1965, BT-Drs. 4/171 in: *Kropff*, AktG 1965, 97; ähnl. *Yi*, Strukturreform 2004, 294: „Der Vorteil des Kollegialprinzips liegt in der Möglichkeit der Selbstkontrolle […] und in der Möglichkeit, Entscheidungen in stärkerem Maße arbeitsteilig zu bewältigen."

[222] *Weber*, Wirtschaft 1922, 673; vgl. auch *Ueberwasser*, Kollegialprinzip 1989, 73: „Dem Kollegium kommt in der Idealvorstellung eine moderierende, machtbrechende Funktion zu: Die Kollegiumsmitglieder sollen sich gegenseitig im Zügel halten."

[223] *Frey*, NZG 2004, 169, 172.

[224] *Dose*, ZGR 1973, 300, 309.

[225] *Fleischer*, NZG 2003, 449, 458 f.; ausf. *von Hein*, ZHR 2002, 464, 496 f.; *Grundei*, BCCG DiskP 2004, 1/12; *Martens* in: Goerdeler u.a., FS Fleck 1988, 191, 199 f.; *Wettich*, Vorstandsorganisation 2008, 38, 300; *Thamm*, Verfassung 2008, 89; *Beckert*, Personalisierte Leitung 2009, 80 bei Fn. 381; *Dauner-Lieb* in: Henssler/Strohn, GftsR 2014, § 78 Rn. 1 a.E.; *Bleicher/Leberl/Paul*, Unternehmensverfassung 1989, 30; *Bainbridge*, Vand L Rev 2002, 1, 55: "internal team governance structures act as a substitute both for external monitoring […] and for legal sanctions".

Dabei kann vorliegend dahinstehen, ob diese Kontrollfunktion den „Vorstand zum ‚Lückenbüßer' des Aufsichtsrats" und zu „seinem eigenen Kontrollorgan" macht, und ob sie „mit der gesetzlichen Rollenverteilung zwischen Vorstand und Aufsichtsrat […] nicht vereinbar" ist.[226]

Zur zweiten Ausprägung als *Organkollektivität* galt die denkbare Alternative – der Generaldirektor an der Spitze der Unternehmenshierarchie – schon dem Gesetzgeber von 1937 nicht sonderlich attraktiv:

„die Schaffung einer Spitze in der Person eines Mannes […] hat, wie die Vergangenheit gelehrt hat, oftmals durch Größenwahn und Verantwortungslosigkeit […] zu schweren Mißbräuchen und zu Zusammenbrüchen von Gesellschaften geführt".[227]

Das griff denn auch der Gesetzgeber von 1965 auf, der das Kollegialprinzip damit begründete, das alternative Generaldirektorenprinzip könne

„für die Gesellschaft gefährlich werden, weil es den Vorsitzenden dazu verleitet, vorschnell und ohne genügende Aussprache mit den übrigen Vorstandsmitgliedern wichtige geschäftliche Entscheidungen zu treffen."[228]

Dem könnte das Kollegialprinzip entgegenwirken, indem es den Entscheidungsprozess abbremst und zugunsten gründlicherer Information entschleunigt. Denn schon Feuerbach konstatierte, dass „was ein Einzelner bloß mit sich selbst abzuthun hat, weit kürzer abgemacht ist, als wenn sich Mehre darüber erst berathen und zu einem gemeinsamen Beschluss vereinigen sollen."[229] Ob die Einführung des Kollegialprinzips zur Entschleunigung der Informationssuche *notwendig* war, lässt sich zwar bezweifeln,[230] vorliegend genügt aber festzustellen, dass sie zu diesem Zweck als geeignet wahrgenommen wird. Denn das Schrifttum geht einhellig davon aus, „dass der Entscheidungsprozess des Kollegialorgans gegenüber dem der monokratischen Organisation langsamer ausfällt",[231] und postuliert mitun-

---

[226] *Bernhardt/Witt*, ZfB 1999, 825, 829, und weiter: die „Zusammenarbeit in einem Kollegialorgan [… sei stets] Wahrnehmung der Leitungsaufgaben, nicht dagegen [.] die Kontrolle der Wahrnehmung dieser Leitungsaufgaben durch den Vorstand anstelle des Aufsichtsrats".

[227] *Schlegelberger/Quassowski u.a.*, AktG 1937, § 70 Rn. 1 (S. 306 a.E.).

[228] Begr. § 77 AktG-RegE 1965, BT-Drs. 4/171 in: *Kropff*, AktG 1965, 99; auch *Oesterle*, zfo 2003, 199, 200 beruft sich auf die Verhinderung „vorschneller" Entscheidungen; *Wettich*, Vorstandsorganisation 2008, 300: „Schutz vor unüberlegten Alleingängen"; *Dauner-Lieb* in: Henssler/Strohn, GftsR 2014, § 78 Rn. 1: „Kollegialprinzip dient dem Schutz der AG vor […] unausgewogenen oder übereilten Handlungen".

[229] *Feuerbach*, Gerechtigkeitspflege 1821, 365.

[230] So *Erle*, AG 1987, 7, 8: Schon das frühere Recht habe Alleinentscheidungen nicht ohne Aussprache zugelassen (m.w.N. in Fn. 12).

[231] *Thamm*, Verfassung 2008, 90; ebenso *Bleicher*, zfbf 1988, 930, 933; *Yi*, Strukturreform 2004, 294; *Seibt* in: Lutter/Schmidt, AktG 2010, § 77 Rn. 2; *Blinder/Morgan*, J Mon Cred Bank 2005, 789, 791: "We have never heard the opposite suggested."; *Brodbeck/Kerschreiter u.a.*, Acad Mgmt Rev 2007, 459; vgl. auch *Schneider*, Beschlussfähigkeit 2000, 42 Fn. 241.

ter sogar einen linearen Zusammenhang: „Je mehr Mitglieder ein Organ hat, desto […] mehr Zeit braucht es für seine Tätigkeiten."[232]

2. Deutsches Verwaltungsorganisationsrecht

Wann Kollegialgremien in der Verwaltung zum Einsatz kommen, richtet sich nach dem Grundsatz der funktionsgerechten Organisationsstruktur, der aus dem Rechtsstaatsprinzip (Art. 20 III GG) folgt.[233] Nach diesem Grundsatz muss sich die Organisation der Verwaltung an ihren Aufgaben ausrichten. Unter ihren Aufgaben lassen sich „routinisierbare, in einem exakten Programm erfassbare Aufgaben und solche mit einem besonderen Bewertungs- und Gestaltungsbedarf" unterscheiden.[234] Letztere seien von „hoher Komplexität" und erforderten daher die „prozedurale Rationalität" eines Kollegialprinzips, das

> „horizontale Strukturen in die Verwaltung ein[führe], die sich nicht nach dem Prinzip von Befehl und Gehorsam richten, sondern auf der Notwendigkeit eigener Meinungsbildung, Diskussion und Überzeugung beruhen."[235]

Dadurch gewährleiste das verwaltungsrechtliche Kollegialprinzip „eine sorgfältige und gründliche Abwägung und Würdigung des Für und Wider und eine, wie man hoffen darf, richtige Entscheidung".[236] Das beruhe auf zweierlei Rationalitätsgewinnen: Erstens werde mehr Sachwissen eingebracht und „unterschiedliche Erfahrungen und Kenntnisse ausgetauscht", weshalb das Kollegialorgan besser in der Lage sei, „neues Wissen zu verarbeiten" und Entscheidungen „tendenziell ausgewogener und sachgerechter" zu fällen.[237] Dem sekundiert das Bundesverfassungsgericht, das in der Kollegialentscheidung „die Vermehrung der entscheidungserheblichen Gesichtspunkte und Argumente, die erhöhte Berücksichtigung von Entscheidungsfolgen und die gesteigerte wechselseitige Kontrolle" angelegt sieht.[238] Zweitens erhöhe das Kollegialprinzip auch die Partizipation: „Wer

---

[232] *Dose*, ZGR 1973, 300, 309.

[233] *Groß*, Kollegialprinzip 1999, 200.

[234] *Groß*, Kollegialprinzip 1999, 204, 207.

[235] *Groß*, Kollegialprinzip 1999, 204, 207, 363 (These 15); ausf. auch *Schneider*, Beschlussfähigkeit 2000, 43 f.

[236] *Dagtoglou*, Kollegialorgane 1960, 22; früher schon *Weber*, Wirtschaft 1922, 163: „Die Kollegialität gewährleistet größere ‚Gründlichkeit' der Erwägungen"; heute ebenso *Groß*, Kollegialprinzip 1999, 205: die „besonders sorgfältige Beratung" biete „größere Richtigkeitsgewähr"; *Jung* in: Hilgendorf/Rudolf, FS Heinz 2012, 883, 891: „Diskurs als Richtigkeitsgewähr".

[237] *Groß*, Kollegialprinzip 1999, 205 f.; *Schneider*, Beschlussfähigkeit 2000, 39; ebenso *Thiele*, Entscheidungsfindung 2008, 63: „sachgerechte Entscheidung durch Beratung".

[238] BVerfGE 91, 148, 166; zust. zit. von *Thiele*, Entscheidungsfindung 2008, 136 Fn. 39 und *Schneider*, Beschlussfähigkeit 2000, 38 Fn. 203 (ebd. 45 näher zum „Kontrollgedanken").

durch die Mitgliedschaft in einem Gremium unmittelbar Einfluss nehmen kann […], bringt sich in der Regel stärker ein als Außenstehende mit nur beratender Funktion."[239]

Diese doppelte „Rationalitätserwartung gegenüber kollegialen Beratungsprozessen" wird nicht auf das öffentliche Organisationsrecht beschränkt, sondern ausdrücklich auch auf „Organisationsformen des Privatrechts" erstreckt.[240] Wenig überraschend beruft sich deshalb auch das gesellschaftsrechtliche Schrifttum auf die verwaltungsorganisatorischen Erörterungen:[241] Zum ersten Rationalitätsgewinn (Zusammenführung von Fachwissen) heißt es etwa: „Es soll erreicht werden, dass die Fähigkeiten, das umfassende (Fach-)Wissen, die größere Erfahrung und die Sichtweisen möglichst vieler verschiedener Personen berücksichtigt werden."[242] Durch die „Ausnutzung des Sachverstandes aller Vorstandsmitglieder"[243] werde „die verfügbare Informationsaufnahme- und -verarbeitungskapazität" erhöht und „das Spektrum der Problemperspektiven bzw. -einschätzungen und somit der Suchraum für mögliche Lösungen erweitert".[244] Das führe im Ergebnis dazu, dass „eine breite informationelle Grundlage und eine gründliche Überlegung der Entscheidung erreicht" werde.[245] Auch der zweite Rationalitätsgewinn (Motivationsgewinn durch Partizipation) findet seinen Widerhall im Gesellschaftsrecht, wenn der „Solidarität der Gruppenmitglieder" bescheinigt wird, sie fördere „die Arbeitsmoral, die wiederum für die Produktivität entscheidend ist",[246] oder wenn umgekehrt formuliert wird, eine „mangelnde Partizipation an der Unternehmensleitung" führe mitunter „zu Demotivationseffekten".[247]

---

[239] *Groß*, Kollegialprinzip 1999, 204; anders akzentuiert *Thiele*, Entscheidungsfindung 2008, 63: Entscheidungen erhielten „einen hohen Grad an Akzeptanz durch die Mitverantwortung der Beteiligten".

[240] *Groß*, Kollegialprinzip 1999, 206, 208 f.

[241] *Wettich*, Vorstandsorganisation 2008 etwa beruft sich vielfach (z.B. 37 Fn. 137, 38 Fn. 44, 303 Fn. 1000) auf *Groß*, Kollegialprinzip 1999 oder (in 38 Fn. 142, 299 Fn. 983, 301 Fn. 989) auf *Dagtoglou*, Kollegialorgane 1960.

[242] *Wettich*, Vorstandsorganisation 2008, 37, 298 f.; *Fleischer*, NZG 2003, 449, 458 f.; *Fleischer* in: Fuchs u.a., FS Immenga 2004, 575, 581.

[243] *Beckert*, Personalisierte Leitung 2009, 80.

[244] *Grundei/von Werder*, AG 2005, 825, 829; *Grundei*, BCCG DiskP 2004, 1/12; *Körner*, ZöR 1937, 55, 62 spricht vom „prozessorganischen Wert" der kollegialen Beratung; *Langevoort*, Geo L J 2001, 797, 810: "Intuitively, the likelihood of multiple viewpoints and different sets of information should be a positive factor in the deliberative process".

[245] *Wettich*, Vorstandsorganisation 2008, 37 f., 299; *Fleischer*, NZG 2003, 449, 458 f.; *Fleischer* in: Fuchs u.a., FS Immenga 2004, 575, 581; *Bleicher/Leberl/Paul*, Unternehmensverfassung 1989, 107; *Dose*, ZGR 1973, 300, 308 spricht pauschal von einem „Leistungsvorteil der miteinander operierenden Mitglieder".

[246] *Dose*, ZGR 1973, 300, 309.

[247] *Grundei*, BCCG DiskP 2004, 1/11 f. m. Verw. auf *Dose*, Vorstandsmitglieder 1975, 70.

3. Deutsches Justizverfassungsrecht

Auch die deutsche Justizverfassung verkörpert eine „im deutschen Rechtskreis seit jeher tief wurzelnde Überzeugung von der Überlegenheit des Kollegialgerichts".[248] Zwar ist die am „Anfang der modernen deutschen Gerichtsverfassung […] fast uneingeschränkte Herrschaft des Kollegialprinzips"[249] inzwischen vielfach durchbrochen, was aber durchweg auf Kostengründe zurückgeführt wird, da „die Vorzüge des Kollegialprinzips nie bezweifelt worden" seien.[250] Die Annahme, „allein das Kollegialgericht biete Gewähr für eine vertrauenswürdige und gerechte Rechtsprechung",[251] stützt das Justizverfassungsrecht wesentlich auf zwei Erwägungen:

Erstens fördere das Kollegialprinzip ein „Klima allseitiger Kritik und damit interner und gegenseitiger Kontrolle", die die Kollegienmitglieder ganz „unbewußt schon durch die bloße Tatsache der Mitberatung und Mitbeschließung übereinander ausüben":[252]

> „Die Vorteile des Kollegialprinzips liegen bekanntlich darin, dass der richterliche Gedankenaustausch und das Ringen um die ‚richtige' Auffassung in einem Kollegialgericht die intensive Durchdringung des Verfahrensstoffs fördern und der juristischen Qualität der Entscheidung zu Gute kommen."[253]

Schon Feuerbach meinte, dass „in der Berathung Meinungen gegen Meinungen in den Kampf" träten,[254] woraus der „Ueberzeugung jedes Einzelnen eine vielseitigere tiefere Ueberlegung, und dem aus diesen Ueberzeugungen hervorgegangenen Gesammtbeschluss die stärkste Vermuthung für

---

[248] *Guillard*, Kollegialgericht 1989, 2 m.Verw. auf die Motive zum GVG (ausf. Zitat 10 f.), weiter 8 und 14 f.

[249] *Ule*, DVBl 1979, 797, 803

[250] *Ule*, DVBl 1979, 797, 804; *Koch/Steinmetz*, DÖV 1981, 50, 53: „allgemeine Ansicht"; *Schmidt* in: Verwaltungsgerichtsordnung 1970, 85, 90: „banal und selbstverständlich"; vgl. auch *Rieß* in: Jung u.a., FS Müller 2008, 599, 609; *Jung* in: Hilgendorf/Rudolf, FS Heinz 2012, 883, 890.

[251] *Guillard*, Kollegialgericht 1989, 2; *Koch/Steinmetz*, DÖV 1981, 50, 53: „Richtigkeitsgewähr"; a.A. *Kissel*, Gerichtsbarkeit 1972, 55 m.w.N.: „Es ist nirgends nachgewiesen, dass vom Einzelrichter […] schlechtere Urteile zu erwarten sind; niemand kann behaupten, dass die Beratung mit zwei anderen Richtern zu einem besseren Ergebnis führen wird".

[252] *Berkemann*, JZ 1971, 537, 540 (erstes Zitat); *Dagtoglou*, Kollegialorgane 1960, 23 (zweites Zitat); *Schmidt* in: Verwaltungsgerichtsordnung 1970, 85, 90; auch *Koch/Steinmetz*, DÖV 1981, 50, 53 nehmen an, „dass die im dauernden Kräftespiel um die Durchsetzung des eigenen Standpunktes sich ausdrückende Meinungs- und Argumentationsvielfalt zu einer kritischen Auseinandersetzung mit der jeweils anderen Meinung zwingt".

[253] Begr. Art. 1 Nr. 3 Justizbeschleunigungsgesetz, BT-Drs. 15/1491 v. 28.8.2003, 17; ähnl. schon Begr. Art. 4 Nr. 5 Entwurf eines Gesetzes zur Entlastung der Rechtspflege, BT-12/1217 v. 27.9.1991, 46: „Das […] Kollegialitätsprinzip ist besonders geeignet, die […] besonders hohe Qualität der getroffenen Entscheidung zu verbürgen."

[254] *Feuerbach*, Gerechtigkeitspflege 1821, 127.

dessen Richtigkeit“ erwachse.[255] Ganz ähnlich beschwor Rudolf von Jhering „den ausgleichenden und zügelnden Einfluss der Collegen“ auf den einzelnen „Gewissenlosen“,[256] erwartete also von der Kollegialentscheidung, dass sie extreme Ansichten oder Ambitionen der Kollegen moderiere und in gemäßigtere Kompromisspositionen überführe. Der kollegialen Organisationsform wird bescheinigt, sie – „und nur sie“ – sei in der Lage, „unterschiedliche subjektive Bewertungen auszugleichen“.[257] Diese vermeintliche „Tendenz der Gruppe, extreme Urteile zu vermeiden und sich auf mittlere Linien zu beschränken“,[258] wurde zur Grundlage ihrer historischen Funktion als „Kompromiss-Kollegialität“, die

> „Delegierte von untereinander kollidierenden ideellen Macht- oder materiellen Interessen zusammenschließen soll[te], um eine *Schlichtung* der Interessengegensätze durch *Kompromiss* zu erreichen“.[259]

Noch heute soll die kollegiale Verwaltung „abgewogene und ausgeglichene“ Entscheidungen[260] und einen „Ausgleich zwischen divergierenden politischen Interessen durch Kompromissbildung“ ermöglichen.[261]

Zweitens wird angenommen, dass das Kollegialprinzip „Einstellungen eines Richters […] zugunsten besser durchdachter und abgeschliffener Erwägungen ‚abzubremsen‘ vermag“.[262] Diese „Langsamkeit, welche die Bedächtlichkeit der Ueberlegung […] verbürgt“, galt schon vor fast zweihun-

[255] *Feuerbach*, Gerechtigkeitspflege 1821, 126.

[256] *von Jhering*, Zweck 1904, 400; zit. von *Guillard*, Kollegialgericht 1989, 9.

[257] *Rieß* in: Jung u.a., FS Müller 2008, 599, 609; so schon *Vierhaus* beim Deutschen Juristentag 1908, zit. in *Schmidt* in: Verwaltungsgerichtsordnung 1970, 85, 90: „Aus dem Zusammenwirken der Richter erscheint […] eine Mannigfaltigkeit gesichert, aus der eine richtige Mittellinie gefunden wird.“; weitere Nachw. aus jener Zeit bei *Haymann* in: Tatarin-Tarnheyden, FS Stammler 1926, 395, 396.

[258] *Berkemann*, JZ 1971, 537, 539, und weiter: „Dieser Kompromisscharakter ist für jeden, der an Gruppenentscheidungen in richterlichen Spruchkörpern teilgenommen hat, augenscheinlich.“; ebenso *Schmidt* in: Verwaltungsgerichtsordnung 1970, 85, 92 „eine Entscheidung nicht der Extreme, sondern der Mitte“; *Rzepka*, BayVBl. 1991, 460, 461: „Neutralisierung vorgefasster subjektiver Einstellungen“; anders offenbar später *Berkemann*, KritV 1988, 29, 48; skeptisch auch schon *Haymann* in: Tatarin-Tarnheyden, FS Stammler 1926, 395, 396 Fn. 3 zur „nicht ganz seltenen“ „Verführung gemäßigter Elemente durch die aufdringliche Rhetorik von Extremen, im Kollegialgericht“.

[259] *Weber*, Wirtschaft 1922, 161; ähnl. ebd. 164 zu „dem Wunsch, spezialistische Fachorientierung und auseinandergehende Interessen sachlicher oder persönlicher Art durch kollegiale Beratung auszugleichen, also: Kompromisse zu ermöglichen.“

[260] Begr. AT VwPO-E, BT-Drs. 10/3437 v. 31.5.1985, 66; wörtlich wiederholt u.a. von *Kopp* in: DJT, Gutachten 1982, B51; *Ule*, DVBl 1979, 797, 804; *Stelkens*, NVwZ 1991, 209, 214; *Hamann*, VerwArch 1992, 201, 202.

[261] *Thiele*, Entscheidungsfindung 2008, 63; *Schneider*, Beschlussfähigkeit 2000, 39 bei Fn. 218.

[262] *Hamann*, VerwArch 1992, 201, 202.

dert Jahren als maßgebliche Rechtfertigung der Kollegialentscheidung.[263] So war es sogar „dem Fürsten bei Einführung des Fachbeamtentums meist nicht unerwünscht", dass mit Kollegialverfahren „fast unvermeidlich eine Hemmung […] vor allem *schneller* Entschließungen" einherging.[264] Dadurch stelle das Kollegialprinzip sicher, dass Entscheidungen „im Ablauf des psychischen Prozesses bei keinem Teilnehmer zu früh kommen".[265]

Obwohl sich das gesellschaftsrechtliche Schrifttum – soweit ersichtlich – nicht auf diese Ausführungen zur Justizorganisation beziehen, kommt es doch zu ähnlichen Schlussfolgerungen: die gemeinsame Beratung im Kollegium ermögliche „ausgewogene"[266] und verhindere „vorschnelle" Entscheidungen (oben bei Fn. 228).

4. Österreichisches Vereins- und Aktienrecht

Die im österreichischen Vereinsrecht 2002 eingeführte Mindestgröße des Vereinsvorstands (oben nach Fn. 132) wurde schon deshalb kritisiert, weil sie noch weiter geht als die insoweit liberalen Organisationsvorgaben des österreichischen Kapitalgesellschaftsrechts.[267] Indessen sollten die strengeren Vorschriften einem „Kontrolldefizit im Bereich der Rechtsform des Vereins" begegnen,[268] dementsprechend führte der Gesetzgeber aus:

> „Mit der vorgeschlagenen Mindestzusammensetzung wird zwar keine zwingende Gesamtgeschäftsführung und Gesamtvertretung, aber doch ein ‚Vier-Augen-Prinzip' eingefordert, ein gewisser Standard gegenseitiger Unterstützung und Kontrolle nach dem Motto ‚vier Augen sehen mehr als zwei'."[269]

Dieses „Vier-Augen-Prinzip", das im Vereinsgesetz nicht allzu konsequent durchgeführt ist,[270] findet sich auch an anderer Stelle im österreichischen Gesellschaftsrecht, denn es wird als maßgeblicher Grund für die spezialgesetzlich angeordnete Mindestbesetzung des Vorstands der Aktiengesellschaft angeführt (vgl. oben bei Fn. 134 und 135).

Für die kollegiale Ausgestaltung des österreichischen Aufsichtsrats hingegen finden sich Zweckerwägungen, die stärker die Bedeutung der gegen-

---

[263] *Feuerbach*, Gerechtigkeitspflege 1821, 365.

[264] *Weber*, Wirtschaft 1922, 162; zit. von *Schneider*, Beschlussfähigkeit 2000, 42 Fn. 239.

[265] *Körner*, ZöR 1937, 55, 60.

[266] *Wettich*, Vorstandsorganisation 2008, 37 f., 298; sprachlich umgekehrt *Oesterle*, zfo 2003, 199, 200: „nicht abgewogene und damit für das Unternehmen gefährliche Entscheidungen" würden verhindert.

[267] Vgl. *Krejci*, VerG 2002, § 5 Rn. 18.

[268] *Krejci*, VerG 2002, § 5 Rn. 18.

[269] So öBReg, Begr. zu § 5 VerG-E, online in *Parlament*, NR: GP XXI RV 990, 2002, S. 26.

[270] Vgl. *Krejci*, VerG 2002, § 5 Rn. 19 f.

seitigen Kooperation als diejenige der gegenseitigen Kontrolle herausstreichen:

„Die Einrichtung als Kollegialorgan soll sicherstellen, dass unterschiedliche Fachkompetenzen und Fähigkeiten zusammenwirken, die Meinungsbildung vielseitig und vielschichtig ist, um für die Entscheidungen eine hohe Richtigkeitsgewähr, jedenfalls ein hohes Argumentationsniveau zu erzielen und zugleich unterschiedliche Interessen im Kontrollgremium zusammenwirken zu lassen.“[271]

5. Schweizerisches Aktienrecht

Das Aktienrecht der Schweiz ist von allen betrachteten Rechtsordnungen die liberalste und kennt nahezu keine gesetzlichen Vorschriften über die Organpluralität. Immerhin aber sieht Ziff. 12 des 2002 erstmals veröffentlichten “Swiss Code of Best Practice” eine entsprechende Empfehlung vor:

„Der Verwaltungsrat soll so klein sein, dass eine effiziente Willensbildung möglich ist, und so gross, dass seine Mitglieder Erfahrung und Wissen aus verschiedenen Bereichen ins Gremium einbringen und die Funktionen von Leitung und Kontrolle unter sich verteilen können. Die Grösse des Gremiums ist auf die Anforderungen des einzelnen Unternehmens abzustimmen.“[272]

Dazu führt der Autor dieses Regelwerks an anderer Stelle weiter aus:

„Je nach Gesellschaftsart und konkreten Gegebenheiten nimmt die *Effizienz des Gremiums* als Informations-, Debattier- und Beschlussorgan von einem gewissen Optimum an mit zunehmender Mitgliederzahl ab. Dies ist eine praktisch unbestrittene Erfahrungstatsache.“[273]

Diese Ausführungen betreffen zwar die Frage, welche Organ*höchst*größe sinnvoll ist,[274] lassen aber Rückschlüsse darauf zu, was Organpluralität und -kollektivität nach Schweizer Lesart bezwecken könnten: Böckli und sein Swiss Code gehen offenkundig davon aus, dass es eine gewisse optimale Organgröße oberhalb von eins gibt.[275] Organe dieser Größe, sollen einerseits „Leitung und Kontrolle“ verteilen, andererseits „Erfahrung und Wissen“ sammeln. Ganz ähnlich heißt es andernorts:

---

[271] *Kalss* in: Doralt/Nowotny/Kalss, öAktG 2012, § 86 Rn. 8.

[272] *Economiesuisse*, Swiss Code 2007, 13 (Ziff. 12 erster Spiegelstr.).

[273] *Böckli*, Schweizer AktR 2009, 1550; ebenso *Roth Pellanda*, Organisation 2007, 209 (Rn. 426): „Ab einer gewissen Grösse wird jede Gruppe durch entstehende Reibungsverluste schwerfällig.“

[274] Selber Blickwinkel bei *Roth Pellanda*, Organisation 2007, 209 Rn. 426: „Bei einer höheren Mitgliederzahl wird die Analyse schwieriger Probleme erschwert; einzelne Beiträge der Mitglieder gehen leichter unter“.

[275] Deutlich *Böckli*, Schweizer AktR 2009, 1550: „Das Optimum liegt [.] doch etwas höher, als manche meinen [… und] ist in vielen Fällen für einen Verwaltungsrat [.] bei 7 bis 10 zu sehen.“

„Grosse Verwaltungsräte ermöglichen ein breites Spektrum an Kenntnissen und Fähigkeiten innerhalb des Gremiums, den Einbezug verschiedener Interessen sowie eine Reduktion der Arbeitsbelastung des einzelnen Mitgliedes.“ – „Jedes Mitglied des Verwaltungsrates bringt seine spezifischen Kenntnisse und Fähigkeiten ein, was einerseits zu mehr Information und besseren Lösungsansätzen und andererseits aber auch zu einer höheren Ausfallsicherheit und einer besseren Möglichkeit zur Elimination von Fehlerquellen führen kann.“[276]

Der Swiss Code betont insoweit, dass es dafür einer „eigenständigen Willensbildung im kritischen Gedankenaustausch mit der Geschäftsleitung“ bedarf,[277] und Böckli wiederum spricht von einem „Kollegium […] in dem eine Funktionszuteilung, eine sinnvolle Debatte und ein Beschluss, der diesen Namen verdient, überhaupt möglich ist.“[278]

6. Andere Rechtsordnungen

Der in *Japan* traditionell etablierte und bis 2006 ausnahmslos vorgeschriebene dreiköpfige Verwaltungsrat wurde ursprünglich

„eingeführt, weil man davon ausging, durch die Diskussion unter den Verwaltungsratsmitgliedern deren Fähigkeiten in einem Gremium zu konzentrieren und dadurch eine optimale Unternehmensführung zu erreichen.“[279]

Ganz ähnlich soll das Kollegialprinzip in *Kanada* durch die Zusammenführung verschiedener Urteilsperspektiven die Kompetenz und Effektivität der Unternehmensführung sicherstellen.[280] Zudem geht das Recht davon aus, dass „kollektive Klugheit und Urteilskraft“ den Schutz der Aktionäre durch eine „gründlichere Überlegung“ geschäftlicher Entscheidungen gewährleisten.[281]

Auch in *England* und *Indien* nutzt das Aktienrecht die „vom Gesetzgeber als äußerst wichtig eingeschätzte Sicherheitsvorkehrung“[282] der kolle-

---

[276] *Roth Pellanda*, Organisation 2007, 211 (Rn. 434) bzw. 209 (Rn. 425).

[277] *Economiesuisse*, Swiss Code 2007, 13 (Ziff. 12 zweiter Spiegelstr.).

[278] *Böckli*, Schweizer AktR 2009, 1732; zust. *Bauen/Venturi*, Verwaltungsrat 2007, 190 (Rn. 505) bei Fn. 200.

[279] *Kawamoto*, Japan 2004, 179 (Rn. 447); wortgleich *Marutschke*, Einführung 2010, 277.

[280] *Iacobucci/Pilkington/Prichard*, Business Corporations 1977, 237: “One means of approaching the objective of securing competent and effective corporate management is […] to designate a minimum number of directors whose judgment must be brought to bear on the business and affairs of the corporation.”

[281] *Iacobucci/Pilkington/Prichard*, Business Corporations 1977, 240: “provides some additional protection to shareholders in that the collective wisdom and judgment of at least three directors will tend to ensure more thorough consideration of management decisions.”

[282] *Palmer/Schmitthoff*, Company Law 1987, 877: “The purpose of these provisions is to maintain a dual control in the management of the affairs of the company and to prevent fraudulent machinations by a one-man management. This purpose is regarded by the legislator as a highly important precaution”.

gialen Entscheidungsfindung, um den möglicherweise „missbräuchlichen Machenschaften eines Einzelvorstands“ vorzubeugen.[283]

Dementsprechend gelangt auch die funktionale Gesellschaftsrechtsvergleichung zu der apodiktischen Schlussfolgerung, dass die plurale Besetzung des Leitungsorgans in erster Linie die „gegenseitige Überwachung erleichtere und eigensinnige Entscheidungen eindämme“.[284]

III. Interdisziplinäre Perspektiven

Schließlich werden die rechtlichen Zweckargumente mit verhaltenswissenschaftlichen Annahmen über die Funktionsweise einer Kollegialentscheidung untermauert:

> „Wichtige Entscheidungen werden eingehender beraten und aus verschiedenen fachlichen Blickwinkeln beurteilt. Außerdem erhöht der Abstimmungsbedarf im Gesamtvorstand den Rechtfertigungsdruck, Beschlussvorlagen besser vorzubereiten und rational zu begründen. Gleichzeitig sinkt die Gefahr vorgefasst-einseitiger Beschlüsse, die man in der empirischen Verhaltensforschung als *bias* bezeichnet.“[285]

Damit sind kognitive Urteilsverzerrungen angesprochen, die im Zusammenhang mit Gruppenentscheidungen und korporativen Akteuren bereits mehrfach untersucht wurden – mit eher gemischtem Befund.[286] Zu denjenigen, deren Abmilderung man sich vom Kollegialprinzip erhofft, gehören vor allem „die Gefahren des *overconfidence* und *self serving bias*“[287], also der kognitiven Urteilsverzerrungen zugunsten von Selbstüberschätzung und Selbstwerterhalt.[288] Genau dasselbe Argument wird im US-amerikanischen Gesellschaftsrecht für die Besetzung des Leitungsorgans mit mehre-

---

[283] *Ramaiya/Chandrachud*, Companies Act 2006, 2592: “The purpose of these provisions is to […] prevent fraudulent machinations by a one-man management.”; vgl. auch vorige Fn.

[284] *Kraakman/Armour u.a.*, Anatomy 2009, 14: “facilitates mutual monitoring and checks idiosyncratic decision-making.”

[285] *Fleischer*, NZG 2003, 449, 458 f. – allerdings bezeichnet *bias* nicht den „vorgefasst-einseitigen Beschluss“, sondern seine kognitive Ursache, vgl. *Fleischer* in: Fuchs u.a., FS Immenga 2004, 575, 581; allg. auch *Wettich*, Vorstandsorganisation 2008, 300; *Grundei/von Werder*, AG 2005, 825, 829; ähnl. *Seibt* in: Lutter/Schmidt, AktG 2010, § 77 Rn. 2.

[286] Übersicht bei *Kerr/MacCoun/Kramer*, Psy Rev 1996, 687, 692 (Tabelle 1); *Baron/Kerr*, Group Process 2003, 131 ff.; *Sniezek*, Org Behav Hum Dec Proc 1992, 124, 135: “information processing in groups can be flawed in the same manner as individual information processing.”; *Engel*, J Inst Econ 2010, 445, 454 f., 459 f., 463: “In many respects, collective and corporate actors suffer from the same biases as individuals.”

[287] *Fleischer* in: Fuchs u.a., FS Immenga 2004, 575, 581; *Fleischer*, NZG 2003, 449, 458 f.

[288] *Fleischer* in: Fuchs u.a., FS Immenga 2004, 575, 581: „*overconfidence bias*, die Neigung von Entscheidungsträgern also, auf ihrer persönlichen Einschätzung zu beharren und die eigenen Fähigkeiten zu überschätzen […] *self serving bias*, wonach Erfolge tendenziell der eigenen Leistung, Misserfolge dagegen den äußeren Umständen zugeordnet werden.“

ren Entscheidungsträgern angeführt.[289] Wiewohl daran mitunter auch Zweifel laut werden,[290] wurden diese bislang weder im US-amerikanischen noch im deutschen Schrifttum aufgegriffen.

IV. Zwischenergebnis: Was verspricht das Kollegialprinzip?

Die Gesamtschau der erörterten Zwecke des Kollegialprinzips ergibt sechs wiederkehrende Argumente. Zunächst diene das Kollegialprinzip einer sinnvollen arbeitsteiligen Bewältigung des Geschäftsanfalls.[291] Allerdings lässt sich dadurch das Kollegialprinzip weder in seiner pluralen noch in seiner kollektiven Ausprägung erklären, denn zum einen ist Arbeitsteilung ebenso gut (womöglich sogar besser) in einer strikten Ressortgliederung möglich, ohne dass es einer kollektiven Entscheidungsfindung bedürfte, zum anderen setzt Arbeitsteilung nicht einmal die plurale Organbesetzung voraus, sondern lässt sich auch in hierarchische Strukturen übersetzen.[292] Damit erscheint die Arbeitsteilung nicht als überzeugendes Argument zur Begründung des Kollegialprinzips. Die anderen Zwecke des Kollegialprinzips beruhen vor allem auf den folgenden fünf empirischen Annahmen:

Motivation durch Partizipation:[293] Das Kollegialprinzip erhöht die Partizipation der Entscheidungsträger und motiviert sie dadurch zur aktiveren Beteiligung an Entscheidungen.

Synergien in der Entscheidungsfindung:[294] Das Kollegialprinzip verbreitert die Informationsgrundlage, indem es verschiedene Informationen und Perspektiven zur Problemlösung zusammenführt.

Entschleunigung der Informationssuche:[295] Das Kollegialprinzip vermeidet vorschnelle Entscheidungen, weil Gruppen im Vorfeld ihrer Entscheidungen sorgfältiger Informationen sammeln.

---

[289] *Bainbridge*, Vand L Rev 2002, 1, 30: "Group decisionmaking presumably checks individual overconfidence by providing critical assessment and alternative viewpoints"; *Dallas*, San Diego L Rev 2003, 781, 785: "group decisionmaking […] decreases […] the chance that a dominant CEO will become convinced of his invincibility".

[290] *Jolls/Sunstein*, J L Stud 2006, 199, 217 f.; *Langevoort* in: Gigerenzer/Engel, Heuristics and Law 2006, 87, 95: "if [a] group is initially populated by overconfident individuals, the resulting dynamic will strengthen the effect, not be a moderating influence."; allg. *Hall*, Manage L 2007, 93, 102: "Just as individuals are subject to cognitive weaknesses in decision making, research suggests that these flaws can be amplified in the group context."

[291] Vgl. z.B. oben bei Fn. 221, 272 und 276.

[292] Vgl. *Groß*, Kollegialprinzip 1999, 204 sowie unten bei und in Fn. 638.

[293] Vgl. z.B. oben bei Fn. 239 und 247.

[294] Vgl. z.B. oben bei Fn. 237–238, 242–245, 271, 276, 279 und 280.

[295] Vgl. z.B. oben bei Fn. 230–229, 262–265 und 281.

Gegenseitige Überwachung und Mäßigung:[296] Das Kollegialprinzip führt durch gegenseitige Kontrolle und Kompromissbildung innerhalb des Organs zu ausgewogenen Entscheidungen.

Verringerung der Selbstüberschätzung:[297] Das Kollegialprinzip reduziert die Selbstüberschätzung, die das kognitive Urteilsvermögen von Entscheidungsträgern systematisch verzerrt.

Diese Liste ist selbstverständlich nicht vollständig, wie das einleitende „insbesondere" verdeutlicht.[298] Dennoch dürfte sie die wichtigsten theoretisch reflektierten und empirisch überprüfbaren Annahmen wiedergeben. Weitere Annahmen beziehen sich auf die *Dynamik* der Gruppenzusammensetzung[299] – etwa „werden sich eher Nachwuchsführungskräfte für die Mitarbeit als gleichberechtigtes Mitglied in einem Kollegium finden"[300] und „sichert das Kollegialprinzip die Kontinuität der Willensbildung und hilft, plötzliche Ausfälle eines Vorstandsmitglieds zu überbrücken."[301] –, sollen hier aber zurückgestellt werden, um zunächst den Fall der *statischen* Gruppe zu verstehen.

Dazu sind im Folgenden die fünf herausgearbeiteten Annahmen anhand der empirischen Forschung zu überprüfen, weil die darin erkennbare

> „Prämisse, die das Kollegialprinzip trägt, dass es nämlich qualitätssteigernd wirkt, nicht hinreichend durch harte Fakten abgesichert ist, sondern eher auf einer tradierten communis opinio […] ruht".[302]

In der deutschen Rechtswissenschaft wurde diese Prämisse bisher nur selten hinterfragt,[303] und auch der US-amerikanischen Rechtswissenschaft wird trotz beachtlicher Vorarbeiten[304] (gerade auch im gesellschaftsrechtli-

---

[296] Vgl. z.B. oben bei Fn. 221–225, 252–261, 268, 269, 283 und 284.

[297] Vgl. z.B. oben bei Fn. 287–290.

[298] Denkbar wäre bspw. auch, Gruppenentscheidungen allein auf ihre politisch legitimierende Wirkung zu stützen, so mit ausf. historischer Begr. *Gevurtz*, Hofstra L Rev 2004, 89.

[299] Ähnl. schon *Berkemann*, KritV 1988, 29, 53 für Richterkollegien.

[300] *Wettich*, Vorstandsorganisation 2008, 300; *Bleicher*, zfbf 1988, 930, 933; *Bleicher/Leberl/Paul*, Unternehmensverfassung 1989, 31; *Rieger* in: Lutter u.a., FS Peltzer 2001, 339, 349; *Yi*, Strukturreform 2004, 294; a.A. für die Justiz *Kissel*, Gerichtsbarkeit 1972, 52 m.w.N.

[301] *Fleischer*, NZG 2003, 449, 458 f.; *Bleicher*, zfbf 1988, 930, 940; *Bleicher/Leberl/Paul*, Unternehmensverfassung 1989, 31, 273; *Wettich*, Vorstandsorganisation 2008, 300; *Beckert*, Personalisierte Leitung 2009, 80; *Seibt* in: Lutter/Schmidt, AktG 2010, § 77 Rn. 2; zum „Kontinuitätsgedanken" im Verwaltungsrecht *Schneider*, Beschlussfähigkeit 2000, 42.

[302] *Jung* in: Hilgendorf/Rudolf, FS Heinz 2012, 883, 890; ebenso *Fanto*, Or L Rev 2004, 435, 467 zur "common sense intuition […], which is enshrined at least in corporate law: that groups enhance the quality of a decision."

[303] Ausn. etwa *Leu/Werner* in: Bender, Tatsachenforschung 1972, 127; *Hendel* in: Bender, Tatsachenforschung 1972, 105; *Hendel*, JZ 1973, 412; *Berkemann*, KritV 1988, 29, 46 ff.

chen Kontext[305]) bescheinigt, dass sie die tatsächliche Funktionsweise von Gruppen bislang kaum erfasst habe.[306]

Bevor das hier versucht wird, ist vorauszuschicken, dass die vorliegende Untersuchung in erster Linie rechtsdogmatisch, nicht rechtspolitisch, motiviert ist. Selbst wenn sich einzelne oder jede der fünf Annahmen als empirisch zweifelhaft herausstellten, wäre damit nicht gesagt, dass das Kollegialprinzip als Institution der Entscheidungsfindung reformbedürftig wäre. Zu dieser und weiteren möglichen Schlussfolgerungen beziehe ich nach der empirischen Sichtung (B. und C.) noch abschließend Stellung (D.).

B. Empirische Forschungsrichtungen und Methoden

Zur Überprüfung der vorgenannten Annahmen können ganz verschiedene empirisch arbeitende Disziplinen beitragen.[307] Umgekehrt wird deshalb eine „zunehmende Fragmentierung des Forschungsgebiets" beklagt,[308] die hier nicht noch dadurch vertieft werden soll, dass die beteiligten Disziplinen trennscharf gegeneinander abgegrenzt werden. Dennoch profitiert die juristische Rezeption von einem Bewusstsein dafür, welcher Wissenschaftstradition und welchen Methoden bestimmte Forschungsrichtungen verpflichtet sind. Die folgende Kategorisierung versteht sich deshalb nur als Überblick, ohne erkenntnistheoretische Abgrenzungsfunktion.

---

[304] Vor allem *Sunstein*, Yale L J 2000, 71; *Sunstein*, J Polit Phil 2002, 175; *Sunstein*, NYU L Rev 2005, 962; *Seidenfeld*, Cornell L Rev 2002, 486, 527 ff.

[305] *Bainbridge*, Vand L Rev 2002, 1; *Dorff*, Cardozo L Rev 2007, 2025; *Pollman*, UC Davis Bus L J 2009, 137; *Sharfman/Toll*, Nw U L Rev Coll 2009, 380.

[306] *Ahdieh*, Chi J Int L 2005, 231: "With a handful of exceptions, however, legal scholars have not focused on groups as concrete, and operational, institutions. […] We have thus failed to attend to questions of how groups actually work."

[307] *Poole/Hollingshead u.a.* in: Poole/Hollingshead, Small Groups 2005, 1: "The study of groups has been an important endeavor in psychology, sociology, education, communication, management, social work, political science, public policy, urban planning, and information science."; ähnl. *Hollingshead/Wittenbaum u.a.* in: Poole/Hollingshead, Small Groups 2005, 21, 25; *Pollman*, UC Davis Bus L J 2009, 137, 165: "Its roots are incredibly multidisciplinary, with researchers from organizational behavior, psychology, operations research, and other disciplines."

[308] *Wittenbaum/Moreland*, Soc Pers Psy Compass 2008, 187, 199, die aber aus der 2006 erfolgten Gründung des "Interdisciplinary Network for Group Research (INGRoup)" neue Hoffnung schöpfen, "that the study of groups will continue to flourish at the crossroads of many fields."; weniger optimistisch noch *Poole/Hollingshead u.a.* in: Poole/Hollingshead, Small Groups 2005, 1, 2: "as a whole, group research remains fragmented and discipline bound. […] Even within fields, lines of research proceed unconnected to other potentially relevant work."

Lässt man die Forschung zum Gruppenverhalten in Tieren außer Betracht – obwohl sie „viele ähnliche Fragen aufwirft wie diejenige zum Menschen“[309] – sind vor allem sechs Forschungsrichtungen auszumachen, die dazu beitragen könnten, die hier entwickelten Grundannahmen zum Kollegialprinzip zu überprüfen. In der Reihenfolge ihrer vermutlichen Relevanz sind das: die rechtsökonomische Boardforschung, die organisationspsychologische Teamforschung, die politikwissenschaftliche Komiteeforschung, die rechtspsychologische Juryforschung, die sozialpsychologische Kleingruppenforschung und die ökonomische Gruppenforschung.[310] Diese sechs Forschungsrichtungen sind zunächst zu skizzieren (I.–VI.) und ihre Tauglichkeit zur Überprüfung der oben genannten Annahmen zu beurteilen (VII.), um nachfolgend den empirischen Erkenntnisstand systematisch erkunden zu können (C.).

I. Der „Board“ in der empirischen Rechtsökonomik

Das Kollegialprinzip ist gesellschaftsrechtlich verankert, daher liegt es zunächst sehr nahe, die empirische Forschung über den Board heranzuziehen, also das Leitungsorgan im monistischen System der US-amerikanischen Unternehmensführung (vgl. schon oben bei Fn. 185).

Die ökonomische Analyse des Rechts – das im anglo-amerikanischen Gesellschaftsrecht vorherrschende Paradigma – betrachtet Gesellschaften als institutionalisierte Vertragsnetzwerke (*nexus of contracts*) zwischen „dem Unternehmen“ einerseits und seinen Produktionsfaktoren andererseits. Damit die Inhaber der Produktionsfaktoren sich nicht gegenseitig übervorteilen, erhalten einige Inhaber des Faktors Kapital keinen festen Profit, sondern nur die Restforderungen der Unternehmenstätigkeit (*residual claims*), was sie dazu anhalten soll, im eigenen Interesse die Inhaber der anderen Faktoren (*agents*) zu überwachen.[311] Der Board bildet in dieser Perspektive den „einheitlichen Gipfel“ des unternehmensinternen Kontrollsystems (*corporate governance*),[312] dient also allein dazu, die Restforderung der Eigenkapitalgeber (Anteilseigner) zu maximieren. Dieses Verständnis lädt empirische Untersuchungen darüber ein, wie sich verschiedene Board-Gestaltungen auf den Wert der Eigenkapitalpositionen auswir-

[309] *Conradt/List*, Phil Trans Royal Soc B 2009, 719: “We might be less aware of it, but group decisions are just as important to social animals as they are for us. […] It emerges that group decisions in animals pose many similar questions to those in humans.” mit nachfolgendem Forschungsbericht; vgl. auch *Couzin/Krause u.a.*, Nat 2005, 513.

[310] Das jeweilige Attribut kennzeichnet nur die Ursprungs- oder Hauptdisziplin der betreffenden Forschungsrichtung, bedeutet dagegen nicht, dass die Forschung ausschließlich in dieser Disziplin angesiedelt wäre.

[311] *Bechtold*, Vertragsrecht 2010, 164 ff.; vgl. auch § 1 Fn. 304.

[312] *Fama/Jensen*, J L Econ 1983, 301, 311, 323.

ken; solche empirischen Untersuchungen werden hier als Boardforschung bezeichnet.[313] Sie entstammt zumeist den Disziplinen der Kapitalmarktforschung[314] sowie der Rechtsökonomik.[315]

Diese Boardforschung nutzt nahezu ausschließlich Korrelationsstudien (dazu § 3 B.II.), was Ursachenrückschlüsse erschwert.[316] Sie korreliert üblicherweise verschiedene Eingabevariablen – wie die Größe, Zusammensetzung oder Beteiligungsverhältnisse des Board[317] – mit dem Börsenkurs. Das bedeutet aber, dass sie sich für das Verhalten *im* Kollegialorgan gar nicht interessiert; begreift man sie überhaupt als *Verhaltens*forschung, dann allenfalls mit Blick auf das Verhalten der Eigenkapitalgeber, das sich im Börsenkurs niederschlägt.[318] Man mag diese Ausrichtung einerseits darauf zurückführen, dass das oben geschilderte Verständnis schon konzeptionell sämtliche Fragen der Unternehmens*führung* letztlich auf solche der Unternehmens*finanzierung* reduziert,[319] oder aber schlicht auf akademische Bequemlichkeit.[320] Gleich aus welchem Grund: Für das Verhalten innerhalb eines Organs ist die Boardforschung nicht empfänglich, Entscheidungs*mechanismen* kann und will sie nicht erklären. „Daher überrascht es nicht, dass man oft hört, dass diese Sichtweise das Unternehmen als *black box* [als undurchschaubaren Prozess] betrachte“.[321]

---

[313] Vgl. *Dallas*, San Diego L Rev 2003, 781, 801 ff. zu weiteren Differenzierungen und anderen Akzentuierungen.

[314] Z.B. *Hermalin/Weisbach*, Fin Mgmt 1991, 101; *Shleifer/Vishny*, J Fin 1997, 737, 751 f.; *Dalton/Daily u.a.*, Strat Mgmt J 1998, 269 (zwei Metastudien über 54 bzw. 31 Primärstudien); *Gillan*, J Corp Fin 2006, 381; *Bebchuk/Cohen/Ferrell*, Rev Fin Stud 2009, 783.

[315] Z.B. *Baysinger/Butler*, J L Econ Org 1985, 101; *Brickley/James*, J L Econ 1987, 161; *Klein*, J Law Econ 1998, 275; *Bhagat/Romano*, Am L Econ Rev 2002, 380, 401 ff.; *Kiel/Nicholson*, Corp Gov 2003, 189; *Williamson*, J L Econ Org 2008, 247; *Adams/Hermalin/Weisbach*, J Econ Lit 2010, 58.

[316] Dementsprechend kommt *Roth Pellanda*, Organisation 2007, 211 (Rn. 430) zu dem Ergebnis, dass solche „Studien nicht zu einheitlichen Ergebnissen kommen“ und „keine allgemein gültigen Empfehlungen“ erlauben.

[317] Detaillierte Beschreibung der verschiedenen Arten von Eingabevariablen bei *Huse*, Brit J Mgmt 2005, S65, S68 f.

[318] *Huse*, Brit J Mgmt 2005, S65, S66: “Actual board behaviour is not explored in these studies, even though some of them use proxies for actual board behaviour.”

[319] Vgl. *Lenoble* in: Cobbaut/Lenoble, Institutionalist Approach 2003, 17, 20: “the question of corporate governance boils down to that of corporate finance”.

[320] *Huse*, Brit J Mgmt 2005, S65, S66: “most research on boards and corporate governance [… was] driven by the ‘publish or perish’ syndrome that is dominating the US academic community. Doctoral students and scholars in tenure track positions have preferred research using easily available data”.

[321] *Hart*, Econ J 1995, 678, und weiter: “that is, the theory predicts how the firm’s production plan varies with input and output prices, but says nothing about how this production plan comes about.”

II. Das „Team“ in der Organisationspsychologie

Der Board ist noch in einer zweiten Forschungsrichtung anzutreffen,[322] nämlich als eine Form dessen, was die Betriebswirtschaftslehre[323] und Organisationspsychologie[324] als Team bezeichnen. Damit sind Gruppen gemeint, die eine gewisse soziale Einbettung und geteilte Identität aufweisen,[325] was allerdings zumeist eher unterstellt als empirisch belegt wird. Kollegialorgane wie der Board gehören jedenfalls zweifelsohne in diese Kategorie.[326]

Aus angewandten Disziplinen kommend, misstraut die Teamforschung vor allem experimentellen Methoden aufgrund ihrer geringen äußeren Gültigkeit, und arbeitet stattdessen eher beobachtend.[327] Den Zugewinn an äußerer Gültigkeit erkauft die Teamforschung allerdings durch ausdrücklichen Verzicht auf eine rigorose Ursachenforschung.[328] Das führt zu mehreren Beschränkungen dieser Forschung, etwa in Bezug auf Teams in der Unternehmensführung (*top management team, TMT*): Zum einen können beobachtende Studien praktisch nie feststellen, inwieweit die Führungsriege im Unternehmen überhaupt kollegial zusammenarbeitet; die höheren Ränge (*upper echelons*) werden schlicht als Team zusammengefasst, obwohl sie vielleicht nur äußerst selten zusammenwirken.[329] Zum anderen verwendet die Teamforschung ein ähnlich grobes analytisches Modell wie die

---

[322] Etwa bei *Pfeffer*, Admin Sci Q 1972, 218; *Payne/Benson/Finegold*, J Mgmt Stud 2009, 704 m.w.N.; *Wagner/Stimpert/Fubara*, J Mgmt Stud 1998, 655 (Metastudie über 29 Primärstudien); konzeptionell übergreifend *Forbes/Milliken*, Acad Mgmt Rev 1999, 489.

[323] Übersichten v.a. im Journal of Management: *Gist/Locke/Taylor*, J Mgmt 1987, 237; *Bettenhausen*, J Mgmt 1991, 345; *Cohen/Bailey*, J Mgmt 1997, 239; *Mathieu/Maynard u.a.*, J Mgmt 2008, 410.

[324] Übersichten etwa bei *Guzzo/Dickson*, Ann Rev Psy 1996, 307; *Ilgen/Hollenbeck u.a.*, Ann Rev Psy 2005, 517; *Kozlowski/Ilgen*, Psy Sci Publ Interest 2006, 77.

[325] *Cohen/Bailey*, J Mgmt 1997, 239, 241: Teams “see themselves and [.] are seen by others as an intact social entity embedded in one or more larger social systems”; *Guzzo/Dickson*, Ann Rev Psy 1996, 307, 309: “groups become teams when they develop a sense of shared commitment and strive for synergy”; ähnl. *Kozlowski/Ilgen*, Psy Sci Publ Interest 2006, 77, 79; weiter *Sutter*, Am Econ Rev 2009, 2247, 2248 Fn. 3: “I will use the term ‘team’ for situations in which several subjects have to agree on a joint decision, and the term ‘group membership’ for situations in which individuals make decisions independently”; noch weiter die Spieltheorie: *Radner*, Ann Math Stat 1962, 857.

[326] *Payne/Benson/Finegold*, J Mgmt Stud 2009, 704, 707 “explicitly consider a corporate board as a team.”

[327] *Guzzo/Dickson*, Ann Rev Psy 1996, 307, 333; *Cohen/Bailey*, J Mgmt 1997, 239, 240.

[328] *Cohen/Bailey*, J Mgmt 1997, 239, 240: “we are willing to sacrifice the rigor of the experimental laboratory to deal with the confounds of the real world in the hope that the findings we identify can be used to guide management practice”.

[329] *Cohen/Bailey*, J Mgmt 1997, 239, 267: „In reality, many of the TMTs […] may not fit our definition of teams, but there is no way to tell“.

Boardforschung, indem sie Eigenschaften des Teams (*input*) mit Prozessen auf Unternehmensebene (*process*) und diese wiederum mit dem Unternehmenserfolg (*output*) korreliert.[330] Dabei soll die Zielgröße „Unternehmenserfolg" (*performance*) zwar ein „objektives Maß für die Leistung der Organisation" bieten,[331] allerdings sind für den Unternehmenserfolg ganz verschiedene Maße denkbar, die untereinander oft kaum korrelieren.[332] Auch die Eingabevariablen sind schwer zu ermitteln, daher beschränken sich viele Studien auf leicht beobachtbare Merkmale wie demographische Eigenschaften der Teammitglieder und Sektorzugehörigkeit des Unternehmens – und vernachlässigen ausgerechnet all jene Kontextfaktoren,[333] die die Wahl der beobachtenden Methode bei erster Annäherung zielführend erscheinen lassen. Schließlich ist der vermittelnde Entscheidungsprozess kaum je direkt zu beobachten, da die Vertraulichkeit von Unternehmensinformationen Feldstudien meist verhindert[334] – und Erfahrungsberichte von Eingeweihten unterbleiben, um die Privatsphäre ihrer Kollegen zu schützen und sich nicht rechtlich angreifbar zu machen.[335]

Im Ergebnis kann also auch die Teamforschung Entscheidungsmechanismen innerhalb von Kollegialorganen nicht durchschauen.[336]

III. Das „Komitee" in der Politikwissenschaft

Auch unter der Bezeichnung als „Komitee" finden sich empirische Studien zum Verhalten von Kollegialgremien. Diese entstammen primär dem expe-

---

[330] Vgl. *Ilgen/Hollenbeck u.a.*, Ann Rev Psy 2005, 517, 519 f.; *Mathieu/Maynard u.a.*, J Mgmt 2008, 410, 412 (je auch zu Erweiterungen des Modells); *Kozlowski/Ilgen*, Psy Sci Publ Interest 2006, 77; *Delarue/Van Hootegem u.a.*, Int J Mgmt Rev's 2008, 127, 131.

[331] *Cohen/Bailey*, J Mgmt 1997, 239, 269.

[332] *Rousseau*, Ann Rev Psy 1997, 515, 525 m.w.N. zu diesem sog. *performance paradox*, das "largely went unnoted"; ähnl. *Mathieu/Maynard u.a.*, J Mgmt 2008, 410, 415: "performance indices are often idiosyncratic"; vgl. schon oben § 3 B.II.2.d).

[333] *Cohen/Bailey*, J Mgmt 1997, 239, 268: "No study collected data on the TMT task. […] Studies also did not consider external communication […] very few considered organizational context factors, except firm size".

[334] *Wheeler* in: Cane/Kritzer, Hdb Empirical 2010, 125, 139 f.: "We should, perhaps, not be surprised by the relative paucity of empirical research by lawyers in this area […] whose inquiries […] require disclosure of market-sensitive information or […] illegal practices. Access to the corporate arena is [.] difficult to obtain if answers are required to questions of a legal nature."

[335] *Payne/Benson/Finegold*, J Mgmt Stud 2009, 704, 705; optimistischer *Huse*, Brit J Mgmt 2005, S65, S75 f.

[336] *Huse*, Brit J Mgmt 2005, S65, S72: "black box of the boardroom"; *Payne/Benson/Finegold*, J Mgmt Stud 2009, 704, 705: "corporate boardroom as a theoretical 'black box'."

rimentellen Zweig der Politologie[337] und der politischen Ökonomik.[338] Genaue Definitionen des „Komitees“ sucht man vergebens, was mit der primär phänomenologischen Entwicklung der Politikwissenschaft zusammenhängen mag. Methodisch dagegen lässt sich die Komiteeforschung genauer bestimmen: Im Gegensatz zu den zuvor betrachteten Forschungsrichtungen verwendet sie die experimentelle Methode und liefert damit Ursachenerklärungen, die den beiden vorgenannten Forschungsrichtungen verborgen bleiben.[339]

Inhaltlich befasst sich die so verstandene Komiteeforschung mit vier Hauptthemen, darunter die Entscheidungsfindung in Komitees und die kollektive Informationsaggregation,[340] die auch vorliegend einschlägig sein könnten. Eine andere Unterteilung identifiziert als gesellschaftsrechtlich relevante Themen der Komiteeforschung vor allem Wahlkampf (*electoral competition*) und Abstimmungsregeln (*voting protocols*) sowie die Thematisierung (*agenda setting*) und Aushandlung (*bargaining*) von Beschlüssen.[341] Allerdings beschränkt sich die Komiteeforschung bei all diesen Themen darauf, die „extrem genauen Hypothesen“ zu überprüfen, die sie aus „komplexen mathematischen Theorien über Komitees, Wahlen und Abstimmungsverhalten“ gewinnt.[342]

Noch erschwert wird die Rezeption dieser Forschungsrichtung dadurch, dass die experimentelle Methode innerhalb der Politikwissenschaft insgesamt noch kaum etabliert ist.[343] So sind zwischen 1926 und 2000 überhaupt nur 105 politikwissenschaftliche Experimente belegt (und nur ein Teil da-

---

[337] Sehr früh etwa *Fiorina/Plott*, Am Polit Sci Rev 1978, 575; *Murnighan/Roth*, J Pers Soc Psy 1980, 92; Übersicht bei *McDermott*, Ann Rev Polit Sci 2002, 31; *McDermott* in: Webster/Sell, Experiments 2007, 483 ff. (Kap. 19).

[338] *Palfrey* in: Wittman/Weingast, Hdb Political Economy 2006, 915; *Palfrey*, Ann Rev Polit Sci 2009, 379.

[339] *Palfrey*, Ann Rev Polit Sci 2009, 379, 380: “permit direct observation of choice behavior […] which in turn permits unusually sharp tests of theoretical predictions”; sehr ausf. *McDermott*, Ann Rev Polit Sci 2002, 31, 33–41.

[340] *Palfrey* in: Wittman/Weingast, Hdb Political Economy 2006, 915, 916: die beiden anderen seien “elections and candidate competition” und “voter turnout and participation games”; vgl. auch *Wilson* in: Webster/Sell, Experiments 2007, 433 ff. (Kap. 17).

[341] Vgl. *Bernhardt/Witt*, ZfB 1999, 825, 836 f. mit dem Fazit: „Vorstandsentscheidungen sind in nicht unbeträchtlichem Umfang abstimmungspolitisch beeinflusste Entscheidungen.“

[342] *Palfrey*, Ann Rev Polit Sci 2009, 379, 380, und weiter: “Most political economy experiments are theory driven. The theoretical framework for these experiments usually comes from noncooperative game theory, cooperative game theory, or the theory of competitive equilibrium in markets.”; beispielhaft *Fiorina/Plott*, Am Polit Sci Rev 1978, 575; *Gerling/Grüner u.a.*, Eur J Polit Econ 2005, 563.

[343] *McDermott*, Ann Rev Polit Sci 2002, 31: “experimentation has been slow to garner a following in political science”; *Palfrey*, Ann Rev Polit Sci 2009, 379, 387: “Laboratory experimentation […] is still a relatively new methodology to political science”.

von zu Komitees), die großteils von sechs oder sieben Forschern stammen und nur zur Hälfte in politikwissenschaftlichen Zeitschriften veröffentlicht wurden.[344] Zudem bleiben politikwissenschaftliche Experimente bislang noch hinter dem methodischen Standard der übrigen experimentellen Sozialwissenschaft zurück.[345] Beides könnte sich zwar in Zukunft ändern, nachdem jüngst eine eigene Fachzeitschrift zur experimentellen politikwissenschaftlichen Forschung ins Leben gerufen wurde.[346] Bisher allerdings existiert zu wenig empirische Substanz, um deren Verlässlichkeit und Aussagekraft halbwegs sicher abzuschätzen.[347]

IV. Die „Jury" in der Rechtspsychologie

Die empirische Forschung zur „Jury" im US-amerikanischen Rechtssystem befasst sich mehr mit einem rechtstatsächlich anzutreffenden Typus als mit einem genau definierten abstrakten Konzept. Es geht ihr um eine Analyse der Wirksamkeit und Wirkungsbedingungen der Geschworenenbank statt um die Analyse allgemeiner Verhaltensmechanismen.[348] Dennoch könnte sie wertvolle Einsichten in das allgemeine Entscheidungsverhalten problemorientierter Beratungsgremien liefern.[349]

Die Juryforschung entstammt einer interdisziplinären Richtung, die sich als Rechtspsychologie (*law and psychology*) oder – in stärkerer Abgrenzung von der Kriminologie – als forensische Psychologie bezeichnen lässt.[350] Nachdem 1955 die wissenschaftliche Beobachtung realer Geschworenenberatungen zu einem öffentlichen Eklat geführt hatte,[351] richtete sich die Juryforschung methodisch primär auf Experimente aus. Zwischen 1955 und 1999 waren 206 empirische Studien zum Verhalten von

---

[344] *McDermott*, Ann Rev Polit Sci 2002, 31, 43 (Forschungsbericht 1926–2000).

[345] *McDermott*, Ann Rev Polit Sci 2002, 31, 44: "the experimental sophistication of typical articles in psychology or economics journals [is] outstripping that of many articles published in political science journals."

[346] *Journal of Experimental Political Science (JEPS)*, gegr. im Februar 2013.

[347] Man vergleiche nur mit der Psychologie, wo einzelne Phänomene oft durch mehrere Metastudien untermauert (z.B. unten Fn. 449) und bisweilen durch Hunderte von Experimenten belegt (z.B. unten Fn. 542) sind.

[348] *Vidmar/Hans*, American Juries 2007, 15: "what juries do and what research tells us about their performance."; *Devine/Clayton u.a.*, Psy Publ Policy L 2001, 622, 712: "improving the jury system".

[349] So der Titel von *Devine/Clayton u.a.*, Psy Publ Policy L 2001, 622: "Empirical Research on Deliberating Groups".

[350] Umfassender Überblick bei *Devine/Clayton u.a.*, Psy Publ Policy L 2001, 622, auf 624 m.w.N. zu früheren Überblicksartikeln.

[351] *Devine/Clayton u.a.*, Psy Publ Policy L 2001, 622, 623: "drew a storm of protest and led the federal government and most states to ban access to the jury room."

und in Juries erschienen, davon 136 Labor- und 13 Feldexperimente.[352] Damit existieren hier zwar mehr Experimente als zu Komitees, dennoch muss im Einzelnen kritisch geprüft werden, zu welchen Themen die Forschung bereits belastbare Schlussfolgerungen erlaubt. Metastudien existieren etwa zum Einfluss, den bestimmte Persönlichkeitsfaktoren,[353] Situationsbedingungen,[354] oder die Gruppengröße[355] auf die Entscheidung der Jury haben.[356] Diese Erkenntnisse werden, soweit möglich, in die nachfolgende Literatursichtung einfließen.

Betrachtet man das von der Juryforschung herangezogene Schrifttum näher, so stellt man allerdings fest, dass diese Forschung zu einem großen Teil auf die allgemeine sozialpsychologische Forschung zum Gruppenverhalten zurückgreift, die ich nun als vorletzte Forschungsrichtung darstelle.

V. Die „Kleingruppe“ in der Sozialpsychologie

Unter einer Gruppe versteht die Sozialpsychologie gemeinhin „zwei oder mehr wechselseitig abhängige Individuen die sich gegenseitig durch soziale Interaktionen beeinflussen“.[357] Diese Definition setzt also keine bestimmte Identität der Gruppe voraus, umfasst also auch Teams (oben II.) jeder Art. Da auch die Teamforschung selten konkretisiert, welche Identität sie einem bestimmten Team unterstellt (vgl. oben nach Fn. 325), ist die Abgrenzung womöglich ohnehin „eher künstlich und sagt mehr über disziplinäres Revierverhalten aus als über grundsätzliche Unterschiede im Hauptaugenmerk oder in der Zielsetzung.“[358] Zumindest im Hinblick auf die disziplinäre Verankerung und die gewählte Methodik ist die Differenzierung hingegen aufschlussreich: Während die Erforschung von „Teams“ zumeist beobachtend in anwendungsbezogenen Disziplinen erfolgt, geht es um „Gruppen“ zumeist in der experimentellen Grundlagenforschung.[359] Dabei spricht man auch von einer *Klein*gruppe, wenn die Gruppe aus höchstens

---

[352] *Devine/Clayton u.a.*, Psy Publ Policy L 2001, 622, 627.

[353] *Sweeney/Haney*, Behav Sci L 1992, 179 (ethnische Zugehörigkeit, 14 Primärstudien); *Narby/Cutler/Moran*, J Appl Psy 1993, 34 (autoritäres Wertesystem, 20 Primärstudien).

[354] *MacCoun/Kerr*, J Pers Soc Psy 1988, 21 (Meinungsasymmetrie, 12 Primärstudien); *Steblay/Besirevic u.a.*, L Hum Behav 1999, 219 (vorprozessuale Berichterstattung, 44 Primärstudien); *Steblay/Hosch u.a.*, L Hum Behav 2006, 469 (unzulässige Beweismittel, 48 Primärstudien).

[355] *Saks/Marti*, L Hum Behav 1997, 451 (17 Primärstudien).

[356] Vgl. auch die Teilliste “Law” bei *Richard/Bond/Stokes-Zoota*, Rev Gen Psy 2003, 331, 357 f.

[357] *Baron/Kerr*, Group Process 2003, 2; *Larson*, Synergy 2010, 20.

[358] *Kerr/Tindale*, Ann Rev Psy 2004, 623, 624; ähnl. *Guzzo/Dickson*, Ann Rev Psy 1996, 307, 309: “degrees of difference, rather than fundamental divergences”.

[359] *Kerr/Tindale*, Ann Rev Psy 2004, 623, 624.

20 Mitgliedern besteht,[360] womit selbst große Kollegialorgane erfasst sind, bis hin zum Aufsichtsrat in Aktiengesellschaften mit mehr als 10 Mio. Euro Grundkapital oder mehr als 20.000 Arbeitnehmern.[361] Damit liegt die Schlussfolgerung nahe, dass Kollegialorgane „den gleichen sozialen und psychologischen Einflüssen unterliegen wie Kleingruppen allgemein".[362]

Die empirische Kleingruppenforschung ist „im Grenzbereich zwischen Soziologie und Sozialpsychologie angesiedelt"[363] und blickt in der Sozialpsychologie auf jahrzehntelange Forschung zurück,[364] wenngleich wohl mit rückläufiger Tendenz seit den 1970er Jahren.[365] Zu vielen Fragen der Sozialpsychologie allgemein und zu Gruppenprozessen insbesondere existieren aussagekräftige Metastudien.[366] In einem jährlich publizierten Überblick über die psychologische Forschung finden sich in den Jahren 1961, 1964, 1967, 1973, 1976, 1979, 1982, 1990 und 2004 ausführliche Forschungsberichte ausschließlich zur Gruppenforschung.[367] Auch andere Periodika enthalten immer wieder Literaturübersichten, aber angesichts der enormen Fülle an Studien beginnt fast jede einschlägige Literaturübersicht mit einer ausdrücklichen Beschränkung ihres Blickfelds.[368] So bleibt Monographien die anspruchsvolle Aufgabe vorbehalten, die zahlreichen und differenzierten Befunde zusammenzuführen und vielversprechende Richtungen für die weitere Forschung zu weisen.[369] Der gewaltige Forschungsapparat hat viel-

---

[360] *Larson*, Synergy 2010, 21, der aber darauf hinweist, dass meist deutlich kleinere Gruppen untersucht werden – "the upper range of what might justifiably qualify as a small group is relatively understudied".

[361] In diesen Fällen hat der Aufsichtsrat nach § 7 I 1 Nr. 3 MitbestG bzw. § 95 S. 4 AktG bis zu 20 bzw. 21 Mitglieder.

[362] *Langevoort*, Geo L J 2001, 797, 810: "Boards of directors are small groups. As such, they are subject to the same social and psychological influences as small groups generally."; ähnl. Berkemann, KritV 1988, 29, 46: kollegiale Spruchpraxis sei „Verhalten ‚im sozialen Kontext' einer Kleingruppe und unterliegt damit den entsprechenden sozialpsychologischen Bedingungen."

[363] *Hendel* in: Bender, Tatsachenforschung 1972, 105, 106 = *Hendel*, JZ 1973, 412.

[364] Zur frühen Forschung 1920–1957 ausf. *Lorge/Fox u.a.*, Psy Bull 1958, 337 m.w.N.

[365] So das Fazit von *Wittenbaum/Moreland*, Soc Pers Psy Compass 2008, 187, 198: "There has been a steady decline of work on intragroup relations, with the numbers reaching a three-decade low in 2006."

[366] Hervorragender Überblick über "474 Meta-Analytic Conclusions" bei *Richard/Bond/Stokes-Zoota*, Rev Gen Psy 2003, 331, insb. 355 f. zu "Group processes".

[367] *McGrath/Kravitz*, Ann Rev Psy 1982, 195 m.w.N.; *Levine/Moreland*, Ann Rev Psy 1990, 585; *Kerr/Tindale*, Ann Rev Psy 2004, 623.

[368] Vgl. nur *Hill*, Psy Bull 1982, 517: "Because this review was intended to be exhaustive, several boundaries were defined."; *Kerr/Tindale*, Ann Rev Psy 2004, 623: "we have omitted many fascinating matters and many of the omissions are intentional."; Engel, J Inst Econ 2010, 445: "There is too much evidence to give a detailed report of each and every study."

[369] Aus der jüngeren Vergangenheit etwa *Baron/Kerr*, Group Process 2003; *Larson*, Synergy 2010; *Laughlin*, Problem Solving 2011.

fältige Erkenntnisse hervorgebracht, dadurch aber das sozialpsychologische Verständnis von Gruppen auch einem permanenten Wandel unterworfen, der die juristische Rezeption erschwert. Einmal rezipierte empirische Befunde müssen deshalb stets neu kritisch geprüft werden.

VI. „Gruppe" und „Team" in der experimentellen Mikroökonomik

Eine letzte empirische Disziplin, die sich mit kollegialem Verhalten befasst, ist der experimentelle Zweig der Mikroökonomik (*experimental economics*). Die traditionelle Volkswirtschaftslehre betrachtet zwar nur den Einzelentscheider als rationalen Maximierer seines eigenen Nutzens, der nicht von seinen Mitmenschen abhängt. In jüngster Zeit allerdings[370] hat auch die ökonomische Forschung ein Interesse an Gruppenentscheidungen entwickelt – beeinflusst maßgeblich durch die Öffnung der ökonomischen Theorie für die Verhaltenswirkungen sozialer Identität.[371]

Die ökonomische Gruppenforschung hat bislang keine eigene Terminologie entwickelt, sondern die Begriffe „Gruppe" und „Team" übernommen, die sie ohne nähere Reflexion synonym verwendet. Nicht zuletzt aufgrund der noch geringen Anzahl der beteiligten Forscher[372] sind ihre Erkenntnisse bislang eher lückenhaft.[373] Forschungssynthesen existieren noch nicht, und auch erste Forschungsberichte[374] wirken eher anekdotisch-illustrativ. Gleichwohl wird als Zwischenergebnis verkündet, dass Gruppen „unterm Strich […] mit höherer Wahrscheinlichkeit Entscheidungen treffen, die den klassischen spieltheoretischen Vorhersagen folgen".[375]

Das Zitat verdeutlicht bereits, dass die ökonomische Gruppenforschung (mehr noch als die politikwissenschaftliche Komiteeforschung) durch ihre spezifische Theorie vorgeprägt ist und ihre Fragestellungen darauf beschränkt. Sie untersucht vor allem Entscheidungen unter Unsicherheit und

---

[370] *Charness/Sutter*, J Econ Persp 2012, 157: "only rather recently […] over the past ten to 15 years".

[371] *Chen/Li*, Am Econ Rev 2009, 431, 432; *Sutter*, Am Econ Rev 2009, 2247: "Largely inspired by George A. Akerlof's and Rachel E. Kranton's (2000) model on the effects of identity on economic outcomes".

[372] Zahlreiche Veröffentlichungen zum Thema stammen von *Gary Charness*, *Martin Kocher* und/oder *Matthias Sutter*.

[373] *Charness/Sutter*, J Econ Persp 2012, 157: "This literature is still young"; *Cheung/Palan*, Exp Econ 2012, 373, 374: "limited economic literature on team decision-making".

[374] *Kugler/Kausel/Kocher*, WIREs Cog Sci 2012, 471; *Charness/Sutter*, J Econ Persp 2012, 157.

[375] *Charness/Sutter*, J Econ Persp 2012, 157, 158; ebenso *Kugler/Kausel/Kocher*, WIREs Cog Sci 2012, 471, 477: "group behavior in games is more in line with rational and selfish predictions than individual behavior is."

strategische Entscheidungen im Sinne der Spieltheorie.[376] Bisher deutet diese Forschung auf drei „Lehren" hin: dass Gruppen kognitiv besser sind, dass sie sich zur Steigerung von Selbstkontrolle und Produktivität eignen, und dass sie insgesamt eigennütziger handeln.[377]

Eine systematische Rezeption der ökonomischen Erkenntnisse ist einstweilen allerdings schwierig. Die ökonomische Gruppenforschung verzichtet weitgehend auf eine Auseinandersetzung mit den anderen oben erwähnten Forschungsrichtungen,[378] und zieht deren Erkenntnisse allenfalls selektiv heran, ohne die einschlägigen Methodenfragen zu reflektieren.[379] Daher harmonieren die ökonomischen Erkenntnisse zum Gruppenverhalten nicht immer mit denen anderer Richtungen und lassen sich nur schwer in einem allgemeineren Zusammenhang würdigen, zumal bislang keine Forschungssynthesen vorliegen, die den robusten Kern der Forschung herausfiltern und in den Kontext der übrigen Forschung einordnen könnten.

VII. Konsequenzen für die vorliegende Untersuchung

Aufgrund ihrer beobachtenden Ausrichtung bleibt sowohl der Board- als auch der Teamforschung das Entscheidungsverhalten *innerhalb* von Kollegialorganen weitgehend verborgen (*black box*)[380]. Dem versuchen die vier experimentellen Forschungsrichtungen abzuhelfen.[381] Unter diesen Richtungen erscheint die Kleingruppenforschung am weitesten entwickelt und am gründlichsten reflektiert, daher wird sie für die vorliegende Untersuchung am ergiebigsten sein.[382] Zwar sind auch einzelne Studien der Komi-

---

[376] *Cheung/Palan*, Exp Econ 2012, 373, 374 f. m.w.N.; ebenso *Kugler/Kausel/Kocher*, WIREs Cog Sci 2012, 471, 473.

[377] *Charness/Sutter*, J Econ Persp 2012, 157, 159 ff., 164 ff., 166 ff.

[378] Für die Psychologie etwa ausdr. *Charness/Sutter*, J Econ Persp 2012, 157, 159 Fn. 5; *Cox/Hayne*, Exp Econ 2006, 209, 210: "Our research involves three important departures from previous work in that we conduct the experiments in an environment in which (a) groups compete in strategic market games, (b) the distinct information possessed by individual group members is varied as an experimental treatment, and (c) the extent to which a group's decisions depart from rationality is quantitatively measured."

[379] Vgl. die Rezeption der Studie von *Michaelsen/Watson/Black*, J Appl Psy 1989, 834 in *Charness/Sutter*, J Econ Persp 2012, 157, 172, die die unten (§ 4 C.II.3.) dargestellte methodische Kontroverse völlig übergeht.

[380] Vgl. oben bei Fn. 321 und in Fn. 336 sowie Ahdieh, Chi J Int L 2005, 231, 236: "Groups remain black boxes in legal analysis, treated as nothing more than aggregations of their individual membership."

[381] Statt aller *Fiorina/Plott*, Am Polit Sci Rev 1978, 575, 576: "Basically, we wish to explain what goes on within the 'black box' of committee decision making."

[382] Auch *Jung* in: Hilgendorf/Rudolf, FS Heinz 2012, 883, 890 empfiehlt die Sozialpsychologie zur Überprüfung des Kollegialprinzips; frühe Rezeptionsversuche bei *Hendel* in: Bender, Tatsachenforschung 1972, 105 und anderen im selben Band, sowie bei *Berkemann*, KritV 1988, 29.

tee- und ökonomischen Gruppenforschung einschlägig,[383] aber eine breiter angelegte Rezeption beider Forschungsrichtungen scheitert daran, dass sie ihren Schwerpunkt auf das strategische Verhalten angesichts von Interessenkonflikten legen,[384] und aufgrund ihres Interesses an formaler Modellierung große Teile des Entscheidungsprozesses ausblenden.[385] Gerade auf jene Teile des Entscheidungsprozesses allerdings – und nicht auf formal modellierbares strategisches Verhalten – beziehen sich die fünf Annahmen des Kollegialprinzips.

Als experimentelle Forschung sieht sich insbesondere die Kleingruppenforschung natürlich dem Einwand ausgesetzt, sie finde „unter künstlichen Bedingungen statt, die mit der Struktur eines Kollegiums kaum vergleichbar sind“ und sei deshalb „auf die Realität außerhalb der Experimentalsituation“ nicht übertragbar.[386] Wie allerdings oben (§ 3 B.III.2.a) dargelegt wurde, setzt Ursachenforschung immer eine gewisse Künstlichkeit und damit einen Mangel an Übertragbarkeit voraus. Das gilt vorliegend umso mehr, da die Annahmen zum Kollegialprinzip gar nicht auf bestimmte Besonderheiten der unternehmerischen Entscheidungsumgebung abstellen, sondern sehr allgemeine Verhaltenshypothesen beinhalten.[387] Dementsprechend führen sie auch für die vorliegenden Fragen durchaus weiter:

> „Es fehlen bislang zwar konkrete Untersuchungen über das gruppendynamische Verhalten eines Aufsichtsrats. Gleichwohl bestehen keine Bedenken, die [experimentellen] Ergebnisse zu verwenden, da sie allgemeine Aussagen der Sozialpsychologie wiedergeben und nicht auf spezielle Gruppen fixiert sind.“[388]

---

[383] Etwa werden die unter C.III. besprochenen Studien nominell zur Komiteeforschung gezählt, vgl. *Gerling/Grüner u.a.*, Eur J Polit Econ 2005, 563, 586 f.

[384] Vgl. *Palfrey* in: Wittman/Weingast, Hdb Political Economy 2006, 915, 917: “Experiments in Committee Decision-Making […] study allocation problems where the members have conflicting preferences over the possible outcomes”; für die ökonomische Forschung oben bei und in Fn. 375 und *Kugler/Kausel/Kocher*, WIREs Cog Sci 2012, 471, 472: “we center our attention on […] comparing individual and group decisions in interactive settings.”

[385] Deutlich *Fiorina/Plott*, Am Polit Sci Rev 1978, 575: “We focus on the period *after* biological, sociological and psychological processes have operated to instill clear preferences in committee members.” (Hervorhebung nur hier)

[386] *Leu/Werner* in: Bender, Tatsachenforschung 1972, 127, 128, und weiter: „Die Messbarkeit setzt hier bestimmte Grenzen.“; ebenso kritisiert *Hendel* in: Bender, Tatsachenforschung 1972, 105, 106, „dass die Thesen der Kleingruppenforschung vorwiegend aus der Beobachtung künstlicher Laborgruppen resultieren. [… Sie] befinden sich auf einem hohen Abstraktionsniveau und berücksichtigen nicht mögliche Abweichungen, die sich aus den Eigentümlichkeiten des besonderen Gegenstandsbereiches ergeben.“

[387] Ähnl. *Fraidin*, UC Davis L Rev 2004, 1, 76: “The relevance of the [empirical group decision-making] research […] to real world settings, such as board meetings, cannot be denied out of hand” und weiter 77: “Although there are differences between laboratory groups and real-world groups, many of these differences are unimportant.”

[388] *Dose*, ZGR 1973, 300, 309.

Auch lassen sich die oben herausgearbeiteten Annahmen nicht durch einfacher übertragbare Beobachtungsstudien erhärten. Praxisbeobachtungen oder Feldstudien können einer Annahme zwar zu höherer Plausibilität verhelfen, können aber nur beschränkt Kausalschlüsse überprüfen und führen mitunter sogar zu erheblichen Fehleinschätzungen der Ursachenzusammenhänge. Insoweit gilt nichts anderes als für die Beurteilung gruppendynamischer Effekt aufgrund anekdotischer Evidenz (Alltagserfahrung):

„Natürlich ermöglicht die Alltagserfahrung wertvolle Einsichten in die Natur und die Konsequenzen kollektiven Verhaltens. Einsichten allerdings, die sich auf informelle Beobachtungen stützen, sind oft recht ungenau oder sogar widersprüchlich. Man betrachte nur paarweise die folgenden Redewendungen:

Zwei Köpfe sind besser als einer, aber *Zu viele Köche verderben den Brei.*
Je mehr, desto besser, aber *Drei sind einer zuviel.* […]
Einigkeit macht stark, aber *Eine Kette ist nur so stark wie ihr schwächstes Glied.*“[389]

Was zunächst paradox erscheinen mag, könnte auf den zweiten Blick kaum näher liegen: Die von der experimentellen Forschung am häufigsten studierten und am gründlichsten durchdrungenen Phänomene sind oft ausgerechnet diejenigen, die den anerkannten Erfahrungssätzen der Praxis zuwiderlaufen.[390] Anschaulich beschreiben etwa die Entdecker der sog. „verborgenen Informationsprofile“ (auf die ich unter C.II.2. eingehe), wie überrascht sie ob ihrer eigenen Entdeckung waren, die jeder damaligen Expertenintuition zuwiderlief – und wie sie selbst und andere Autoren in zahlreichen Studien versuchten, den neuen Befund zu widerlegen oder in das alte Schema zu pressen, bevor sie ihre eigene Entdeckung ernst nahmen und deren Tragweite begriffen.[391] Gruppendynamiken sind oft latent, so dass sogar der Bundesgerichtshof 1954 feststellte, es liege

„im Wesen einer gemeinschaftlichen Willensbildung mehrerer Personen, dass sich in der Regel die Willensbildungen der einzelnen Teilnehmer […] wechselseitig zu beeinflussen pflegen. […] Diese sich im psychischen Bereich vollziehenden Vorgänge werden häufig den Beteiligten selbst gar nicht bewusst und sind deshalb ihrer Natur nach einer nachträglichen Erfassung […] nur schwer zugänglich.“[392]

Daher mögen Praxisbeobachtung und gesunder Menschenverstand zwar notwendige Bedingungen für die Beurteilung von Gruppenprozessen sein, allein ausreichend sind sie aber nicht (vgl. auch oben § 2 A.I.1.).

---

[389] *Steiner*, Group Productivity 1972, 3 f.; *Hendel* in: Bender, Tatsachenforschung 1972, 105: „Daß die allgemeine Alltagserfahrung keine eindeutigen Antworten parat hat, zeigen schon die widersprüchlichen Sprichwörterweisheiten wie: ‚Viele Köche verderben den Brei‘ oder: ‚So viel Köpfe, so viel Sinn‘ und im Gegensatz dazu: ‚6 Augen sehen mehr als 2‘.“

[390] Vgl. schon § 2 Fn. 26.

[391] *Stasser/Titus*, Psy Inq 2003, 304, 305 f.

[392] BGHZ 12, 327, 331.

Für die vorliegende Untersuchung bedeutet das, dass Erkenntnisse aus der beobachtenden Forschung sowie allfällige anekdotische Evidenz zwar zur Überprüfung oder Illustration einzelner Schlussfolgerungen in die Darstellung eingehen werden, dass diese Schlussfolgerungen allerdings hauptsächlich auf experimentellen Studien der kausalen Wirkungsmechanismen beruhen werden. Nur diese können letztlich die Art von Annahmen untermauern oder erschüttern, die das Gesellschaftsrecht über das Kollegialprinzip macht.

C. Würdigung der empirischen Erkenntnisse

In der systematischen Sichtung der empirischen Literatur sehe ich mich vor allem der oben (§ 2 B.VII.) entwickelten Erkenntnishierarchie von Forschungsauswertung, Forschungsbericht und Primärstudie verpflichtet, werde also in erster Linie Metastudien heranziehen und Forschungsberichte sowie Primärstudien nur dort erörtern, wo es an Metastudien fehlt. Dieses Vorgehen ist gerade auch für die hier zu behandelnden Fragen der Gruppenforschung zielführend.[393]

Da die empirische Gruppenforschung allerdings sogar mit Rücksicht auf diese methodische Beschränkung einen kaum überschaubaren Umfang erreicht hat, werde ich die Darstellung konsequent problembezogen ausrichten. Die Sichtung und Darstellung folgt also der Reihenfolge der oben entwickelten fünf Grundannahmen des Kollegialprinzips. Dabei müssen viele interessante und praktisch relevante Befunde unberücksichtigt bleiben.[394] Auch werde ich im Sinne analytischer Klarheit verschiedene Phänomene, die in der Realität untereinander verbunden sind, so genau wie möglich gegeneinander abgrenzen und Querverbindungen einstweilen ausblenden. Zur Synthese und Bewertung dieser Phänomene nimmt erst der nächste Abschnitt (D.I.) Stellung.

I. Motivation durch Partizipation?

Die erste Annahme zum gesellschaftsrechtlichen Kollegialprinzip lautet, dass es Entscheidungsträger zur aktiveren Beteiligung an Entscheidungen

---

[393] *Hollingshead/Wittenbaum u.a.* in: Poole/Hollingshead, Small Groups 2005, 21, 50: "Meta-analyses can find patterns across the primary studies that will allow researchers to better identify and apply the fundamental principles of small groups".

[394] Etwa zur optimalen Gruppengröße oder Führungsstruktur, zum sozialen Konflikt innerhalb und zwischen Gruppen, zur Entscheidungslegitimation durch Partizipation, usw. Im Übrigen untersucht die empirische Forschung auch nicht „das" Kollegialprinzip, sondern nur den Überschneidungsbereich von faktischer Organpluralität und Organkollektivität – also tatsächlich gleichberechtigte gemeinsame Entscheidungen mehrerer.

motiviert (vgl. oben bei Fn. 293). Demgegenüber erscheint es aber ebenso denkbar – und nicht weniger plausibel – dass „der Wegfall des Ansporns, der in der *Allein*verantwortlichkeit liegt, das Interesse an energischer Mitarbeit mindern“ könnte,[395] weil „sich der einzelne als anonymes Glied des Kollektivs fühlt und damit sein Verantwortlichkeitsgefühl schwindet“.[396] Das Kollegialprinzip könnte also „zu einer Verwässerung der Verantwortlichkeiten und einem Nachlassen der individuellen Arbeitsleistung“ führen, „während die direkt zurechenbare Letztverantwortung zu vergleichsweise großer Leistungsbereitschaft anspornt.“[397]

Wie sich die Zugehörigkeit zu einer Gruppe auf die Leistungsbereitschaft auswirkt, war die erste Fragestellung überhaupt, zu der sozialpsychologische Experimente durchgeführt wurden.[398] Die ersten dieser Experimente gehen zurück auf den französischen Agraringenieur Max Ringelmann und wurden 1913 veröffentlicht, aber nach längerer Verschollenheit erst 1986 wiederentdeckt.[399] Ringelmann maß den individuellen Einsatz von Gruppenmitgliedern unter anderem beim Tauziehen und stellte fest, dass größere Gruppen zwar in der Summe eine höhere, im Durchschnitt aber eine niedrigere Zugkraft entfalteten als kleinere Gruppen: Während die Mitglieder von Zweiergruppen durchschnittlich 93 % ihres individuellen Potentials beitrugen, fiel diese Rate auf 85 % in Dreiergruppen – und in noch größeren Gruppen fast linear, bis auf nur noch 49 % in Achtergruppen.[400] In einer Gruppe mit sieben Gleichgesinnten steuerte der Einzelne also weniger als die Hälfte seines Potentials bei – und das achte Gruppen-

---

[395] *Körner*, ZöR 1937, 55, 55 f. hielt das für eine „kleinlich-eitle Erwägung“, die einem „halbwegs charaktervollen Mensch“ fernliege; *Dagtoglou*, Kollegialorgane 1960, 25 dagegen beruft sich auf „die dem Menschen angeborene Neigung zur *Flucht vor persönlicher Verantwortung*“; im gesellschaftsrechtlichen Schrifttum ebenso *Thamm*, Verfassung 2008, 91: „Flucht vor der persönlichen Verantwortung in die Anonymität des Kollegialorgans“.

[396] *Wettich*, Vorstandsorganisation 2008, 300; ebenso *Schmidt* in: Verwaltungsgerichtsordnung 1970, 85, 94: „Minderung der Verantwortungsfreude der Beisitzer“; *Berkemann*, JZ 1971, 537, 540; schon laut *Weber*, Wirtschaft 1922, 163 „*teilt* [Kollegialität] die Verantwortlichkeit, und bei größeren Gremien schwindet diese gänzlich“; *Roth Pellanda*, Organisation 2007, 209 (Rn. 425) Fn. 1383: „Verwischung und Verteilung der Verantwortlichkeiten“; ähnl. *Bleicher*, zfbf 1988, 930, 933; *Yi*, Strukturreform 2004, 294 bei Fn. 1600.

[397] *Fleischer*, NZG 2003, 449, 458; *Grundei*, BCCG DiskP 2004, 1/11; *Weibel/Rost/Osterloh*, zfbf 2007, 1029, 1033 f.: „Komplexe, wissensintensive Tätigkeiten im Team erfordern […] ein hohes Ausmaß an intrinsischer Motivation, weil ihre Ergebnisse nicht beobachtet und zugerechnet werden können.“

[398] *Bond/Titus*, Psy Bull 1983, 265, 284 zitieren *Gordon Allport*, das sei “the first experimental problem – and indeed the only problem for the first three decades of experimental research” gewesen.

[399] Wiederentdeckt durch *Kravitz/Martin*, J Pers Soc Psy 1986, 936, durchgeführt bereits 1882–1887 (ebd. 937); damit gebührt ihm “the distinction of being the first experiment in social psychology”, *Baron/Kerr*, Group Process 2003, 47.

mitglied erhöhte die Gesamtkraft der Gruppe gar nicht mehr, weil sein Einsatz von der Einsatzminderung aller anderen aufgewogen wurde.[401] Die genauen Daten dieses Experiments sind etwas fragwürdig,[402] doch der Ringelmann-Effekt gilt als wesentlicher Grund für die Befürchtung, dass auch Kollegialentscheidungen im Gesellschaftsrecht die individuelle Leistungsmotivation vermindern könnten.[403]

Einige Jahre später als Ringelmann belegte indessen Otto Köhler 1926 und 1927 einen genau gegenläufigen Effekt, der ebenfalls zunächst in Vergessenheit geriet und erst Ende der 1980er Jahre wiederentdeckt und nach seinem Erstentdecker benannt wurde.[404] Köhler ließ seine Versuchsteilnehmer gemeinsam eine Stange stemmen, die mit 41 kg Gewicht je Gruppenmitglied beschwert war (also mit 82 kg in der Zweier- und 123 kg in der Dreiergruppe) und verglich die Gruppenleistung mit der Leistung des Einzelnen am 41-kg-Gewicht. Dabei beobachtete er, dass Gruppen mit ungefähr gleich starken Mitgliedern sowie Gruppen mit ganz ungleich starken Mitgliedern zwar einen deutlichen Leistungsabfall verzeichneten (auf 83,9 % der durchschnittlichen Einzelleistungen in Zweier- und 74,4 % in Dreiergruppen), dass Gruppen hingegen, deren Mitglieder einen *mittleren* Kraftunterschied aufwiesen, den Durchschnitt ihrer Mitglieder sogar um 15 % in Zweier- und 5 % in Dreiergruppen *über*trafen.[405] Den mit 35 % größten Leistungszuwachs erfuhren dabei Gruppen, deren schwächeres Mitglied etwa drei Viertel der Kraft des stärkeren besaß. Das könnte bedeuten, dass Gruppenzugehörigkeit zu höherer Einsatzbereitschaft anspornen *kann*, wenn und weil sie einen als überwindbar wahrgenommenen Fähigkeitsunterschied zwischen den Mitgliedern etabliert.

Der Ringelmann- und der Köhler-Effekt veranschaulichen zwei Phänomene, die in der Forschung zur Gruppenmotivation unter den Oberbegrif-

---

[400] Kravitz/Martin, J Pers Soc Psy 1986, 936, 938 Tab. 2; illustriert in *Drewes/Schultze/Schulz-Hardt* in: Frey/Bierhoff, Interaktion und Gruppe 2011, 221, 231; *Bainbridge*, Vand L Rev 2002, 1, 11.

[401] Vgl. *Kravitz/Martin*, J Pers Soc Psy 1986, 936, 938 Tab. 2: 7 Mitglieder á 56 % = 8 Mitglieder á 49 % = 392 %.

[402] Krit. zur Tauziehlegende *Kravitz/Martin*, J Pers Soc Psy 1986, 936, 938: "Ringelmann did not specify the tasks on which the data in Table 2 are based. [… He] provided less information about these data than about any other data discussed in his article. It is ironic that it has been precisely these data that have had such a profound impact on social psychology."; ebenso *Witte*, Eur J Soc Psy 1989, 147, 148: "data given by Ringelmann [.] were drawn from a number of uncontrolled field studies, and should not be taken too seriously."

[403] *Bainbridge*, Vand L Rev 2002, 1, 11: "While board decisionmaking differs rather dramatically from tug-of-war, members of a multimember board likely engage in a certain amount of social loafing."

[404] *Witte*, Eur J Soc Psy 1989, 147.

[405] *Witte*, Eur J Soc Psy 1989, 147, 149 ff.; *Larson*, Synergy 2010, 284 ff.

fen soziale Hemmung (*social inhibition*) und soziale Erleichterung (*social facilitation*) geführt werden.[406] Drei verschiedene Metastudien haben im Verlauf der letzten dreißig Jahre wichtige Erkenntnisse dieser Forschung herausdestilliert.

Die erste Metastudie untersuchte alle 202 Primärstudien zu sozialen Einflüssen auf die menschliche Motivation, die vor 1982 in englischer Sprache publizierten worden waren, sowie 39 unveröffentlichte Studien.[407] Sie kam zu dem Ergebnis, dass es vor allem von der Art der Aufgabenstellung abhänge, ob Gruppenarbeit zu sozialer Hemmung oder zu sozialer Erleichterung führe: Die Lösung komplexer Aufgaben werde gehemmt, die Lösung einfacher Aufgaben hingegen erleichtert.[408] Beide Effekte seien allerdings überraschend gering.[409]

Die zweite Metastudie untersuchte genau achtzig Jahre nach der Veröffentlichung *Ringelmanns* 78 Primärstudien zur sozialen Hemmung, die zwischen 1974 und 1991 erschienen waren.[410] Zur Aufgabenkomplexität waren in jener Metastudie kaum verlässliche Aussagen möglich,[411] allerdings konnte die Studie aufzeigen, dass soziale Hemmung sowohl in Laborexperimenten als auch in Feldstudien robust alle Arten von Aufgaben betrifft und bei allen Arten von Versuchseilnehmern auftritt – wenn auch in Feldstudien, bei Frauen und bei Angehörigen östlicher Kulturen etwas geringer, so dennoch systematisch nachweisbar.[412] Die Metastudie untersuchte eine Reihe von Einflussfaktoren, und erklärte soziale Hemmung letztlich vor allem als Resultat doppelter Unsicherheit des Gruppenmitglieds: einerseits darüber, ob sein Einsatz zum gewünschten Ergebnis führe, andererseits darüber, ob ein bestimmtes Ergebnis erwünscht sei.[413]

---

[406] *Baron/Kerr*, Group Process 2003, 20 ff.; *Larson*, Synergy 2010, 259 ff.; je m.w.N.; im juristischen Schrifttum erwähnt nur *Berkemann*, KritV 1988, 29, 46 die *social facilitation*.

[407] *Bond/Titus*, Psy Bull 1983, 265, 267 f. m.w.N. zu früheren Literaturübersichten.

[408] *Bond/Titus*, Psy Bull 1983, 265, 274: d = -0,36 bzw. d = 0,11, jeweils p < 0,005.

[409] *Bond/Titus*, Psy Bull 1983, 265, 272: "Even this largest [inhibitory] effect accounts for only 3.1% of the between-subjects variance in performance quality." und 274: "the facilitatory mean effect is quite small; on the average, it accounts for only .30% of the performance variance between subjects."

[410] *Karau/Williams*, J Pers Soc Psy 1993, 681 (auf 682 findet sich eine "Brief History" der Forschung); neuere Studien nennen *Baron/Kerr*, Group Process 2003, 53.

[411] *Karau/Williams*, J Pers Soc Psy 1993, 681, 697 Fn. 7: "only seven comparisons involving complex tasks were available […] Moreover, four of the studies coded as using a complex task […] may not have been difficult enough".

[412] *Karau/Williams*, J Pers Soc Psy 1993, 681, 700: "although the magnitude of the effect was reduced for field studies, for women, and for subjects in Eastern cultures, social loafing was still significant under all of these conditions."

[413] *Karau/Williams*, J Pers Soc Psy 1993, 681, 700: "individuals' perceptions of either the instrumentality of their efforts for obtaining valued outcomes or the degree to which outcomes in a particular setting are likely to be valued."

Aus dieser Differenzierung folgen zwei verschiedene Perspektiven, die im gesellschaftsrechtlichen Schrifttum bisher nicht genau abgegrenzt werden:[414] Soweit das Gruppenmitglied nicht weiß, ob sein Beitrag zum gewünschten Ergebnis führt, könnte es ihn als entbehrlich empfinden (*high dispensability*) und sich deshalb als Trittbrettfahrer verhalten (*free riding*).[415] Soweit es hingegen seinen Beitrag nicht als entbehrlich, wohl aber als nicht abgrenzbar und für Dritte damit nicht bewertbar hält (*low identifiability*),[416] könnte es auf Kosten der übrigen Gruppenmitglieder faulenzen (*social loafing*).[417] Das sind die beiden Hauptausdrucksformen der sozialen Hemmung.

Die dritte Metastudie schließlich widmete sich der sozialen Erleichterung und untersuchte 23 Primärstudien zwischen 2000 und 2007.[418] Diese vergleichsweise geringe Anzahl und das junge Datum der Studien dürften darauf zurückzuführen sein, dass die Gruppenforschung aus historischen Gründen auf die Nachteile von Gruppenentscheidungen (*process losses*) fixiert war und erst vor kurzem auch deren Vorteile in den Blick genommen hat.[419] Die Metastudie zur sozialen Erleichterung nun kam zu dem Ergebnis, dass schwächere Gruppenmitglieder durch die gemeinsame Arbeit mit Überlegenen durchaus zu höherer Leistung angeregt werden.[420] Das wird auf zwei Prozesse zurückgeführt: einerseits können sich schwächere Gruppenmitglieder nur mit besseren vergleichen (*upward social comparison*),

---

[414] *Fanto*, Or L Rev 2004, 435, 467 Fn. 106 setzt sie gleich; *Sibert*, Int Fin 2006, 145, 147 meint, dass *free riding* in der Ökonomik mit *social loafing* in der Sozialpsychologie identisch sei; *Bainbridge*, Vand L Rev 2002, 1, 40 f. unterscheidet zwar "social loafing" von "collective action failures" in der Form des "potential free riding problem", aber ohne klares Differenzierungskriterium; sehr präzise definiert *Thamm*, Verfassung 2008, 91 das soziale Faulenzen, geht aber nicht auf das Trittbrettfahren ein.

[415] *Larson*, Synergy 2010, 276 ff., 279 f.; vgl. auch *Baron/Kerr*, Group Process 2003, 56 f.

[416] Dieses Problem der nicht abgrenzbaren Verursachungsbeiträge begegnet außerhalb von Gruppenentscheidungen etwa in der Dogmatik zur alternativen Kausalität (§§ 830 I 2 BGB, 231 StGB), dazu schon § 3 bei Fn. 19.

[417] *Larson*, Synergy 2010, 274 ff., 279 f.; vgl. auch *Bainbridge*, Vand L Rev 2002, 1, 10 f. sowie *Baron/Kerr*, Group Process 2003, 55 mit dem Hinweis auf einige Studien zweier Autoren, nach denen allein die mögliche Bewertung der Gruppenleistung durch Dritte, oder gar durch das Gruppenmitglied selbst, soziales Faulenzen reduziere: "the original loafing effect appears to be restricted to a fairly narrow range of group task settings where evaluation of member or group performance by *anyone* is unlikely." (Hervorhebung nur hier).

[418] *Weber/Hertel*, J Pers Soc Psy 2007, 973, auf 989 kritisch zur Unschärfe des Begriffs "social facilitation".

[419] Näher zur Entwicklung *Larson*, Synergy 2010, 18 f.: die Prozessverlustlehre von *Ivan Steiner* „has promoted a research literature that has become decidedly unbalanced“; früher schon *Hackman/Morris*, Adv Exp Soc Psy 1975, 45, 47.

[420] *Weber/Hertel*, J Pers Soc Psy 2007, 973, 979: "significant and moderate overall effort increase of individuals working either as a group or coactively with a superior partner (g = . 60; d = .68; k = 71; CI = .53, .66)."

ohne dass das viel mit der Gruppenarbeit als solcher zu tun habe, andererseits empfänden schwächere Gruppenmitglieder ihre Bemühungen mitunter als unverzichtbar für den Gruppenerfolg (*indispensability*).[421] Dieser zweite Prozess setze Aufgaben voraus, bei denen die Gruppenleistung von derjenigen des schwächsten Mitglieds maßgeblich abhängt – und erklärt damit auch den Köhler-Effekt beim Gewichtheben, wo das erste Mitglied, das aufgibt, die Bemühungen der Gruppe insgesamt beendet.[422] Auch stellte die Metastudie fest, dass der motivierende Effekt der Gruppenarbeit stärker bei körperlichen als bei geistigen Aufgaben auftritt.[423]

Der empirische Erkenntnisstand zur Auswirkung von Gruppenarbeit auf die Leistungsbereitschaft ist damit reich an Schattierungen.[424] Ob die Gruppe den Einzelnen zu mehr oder weniger Einsatz anregt, hängt maßgeblich von der Art und Komplexität der Aufgabe, von der Identifizierbarkeit der Einzelleistung und davon ab, inwieweit einzelne Leistungen entbehrlich oder unverzichtbar sind.

II. Synergien in der Entscheidungsfindung?

Die zweite Annahme zum gesellschaftsrechtlichen Kollegialprinzip lautet, dass es die Informationsgrundlage für Entscheidungen vergrößert, indem es verschiedene Informationen und Perspektiven zusammenführt (vgl. oben bei Fn. 294). Ob und inwieweit das Gruppen tatsächlich gelingt, gehört zu den traditionellen Grundproblemen der empirischen Verhaltensforschung. Leider arbeitet diese Forschung – nicht zuletzt aufgrund ihrer Zersplitterung in verschiedene Strömungen – mit unterschiedlichen Begriffen, die nicht immer sauber genug differenziert werden.[425] Das führt gelegentlich zu erheblichen Missverständnissen in der Formulierung von Forschungsfragen und Auswertung von Ergebnissen, die in der juristischen Rezeption nicht immer erkannt werden.[426] Daher kläre ich zunächst die Grundkonzepte der empirischen Forschung zur Entscheidungsfindung in Gruppen.

---

[421] *Weber/Hertel*, J Pers Soc Psy 2007, 973, 986.

[422] Aufgaben dieser Art werden mit der *Steiner*schen Typologie (dazu sogleich bei Fn. 428) als konjunktiv bezeichnet.

[423] *Weber/Hertel*, J Pers Soc Psy 2007, 973, 985: “Motivation gains were significantly higher for physical tasks than for cognitive tasks ($g = .82$; $k = 32$; CI = .72, .91, vs. $g = .38$; $k = 39$; CI = .28, .47; $p_{diff} < .001$).”

[424] Vgl. *Weber/Hertel*, J Pers Soc Psy 2007, 973, 989: “Collective work [.] extends both negative and positive action consequences of an individual.”

[425] So auch *Hill*, Psy Bull 1982, 517, 519: “The need for precise terminology can be seen in the history of individual versus group comparisons.”

[426] Das beste Beispiel bietet die juristische Rezeption der vielzitierten Studie von *Michaelsen/Watson/Black*, J Appl Psy 1989, 834, dazu unten bei Fn. 478.

1. Konzeptionelle Vorabklärungen

Ausgangspunkt dieser Klärung ist die Feststellung, dass der größte Teil der empirischen Gruppenforschung die Hypothese untersucht, dass Aufgaben in der Gruppe besser bewältigt werden als einzeln.[427] Aus dieser Feststellung ergeben sich drei konzeptionelle Fragen: Was ist eine Aufgabe? Wann ist sie besser bewältigt? Im Vergleich zu welchem Einzelnen?

Aufgabenarten. Völlig einmütig beruft sich die heutige Gruppenforschung auf eine 1972 erschienene Abhandlung von Ivan Steiner, die den damaligen Erkenntnisstand der sozialpsychologischen Gruppenforschung nach den Aufgabenarten klassifizierte, die einer Gruppe begegnen können.[428] Diese Klassifikation beinhaltet fünf Kategorien von Aufgaben: teilbare, additive, disjunktive, konjunktive und diskretionäre.[429] Ohne alle diese Aufgabenarten näher darzustellen, genügt hier die Feststellung, dass Aufgaben, die eine Auswahl unter mehreren Alternativen erfordern – also *Entscheidung*saufgaben, wie sie sich Kollegialorganen regelmäßig stellen – als disjunktiv bezeichnet werden. Dementsprechend beschränkt sich die nachfolgende Darstellung weitestgehend auf solche disjunktiven Aufgaben.

Qualitätsmaßstab. Eine weitere klassische Differenzierung von Gruppenaufgaben führte Patrick Laughlin 1980 in die Diskussion ein.[430] Er postulierte ein Kontinuum von Aufgaben zwischen reinen Intelligenzaufgaben, die eine logisch zweifelsfrei beweisbare Lösung haben (*intellective tasks*), und reinen Beurteilungsaufgaben, die ein normatives Werturteil erfordern (*judgmental tasks*).[431] Da reine Beurteilungsaufgaben nicht objektiv „besser“ oder „schlechter“ gelöst werden können, fallen sie grundsätzlich aus der empirischen Forschung zur Entscheidungsqualität heraus.[432] Bei ihren Studien legt die Forschung also Entscheidungsaufgaben zugrunde, die nach einem bestimmten Kriterium als „richtig oder falsch“, „besser oder schlechter“ gekennzeichnet werden können. Das ruft immer wieder die Kritik auf den Plan, dass die Gruppenforschung gerade deshalb für Unter-

---

[427] Ähnl. *Leu/Werner* in: Bender, Tatsachenforschung 1972, 127: „Der Vergleich des Einzelnen mit der Gruppe unter dem Gesichtspunkt ihrer Leistungsfähigkeit ist ein geradezu klassisches Thema der experimentellen Sozialpsychologie.“

[428] *Steiner*, Group Productivity 1972, 14 ff. (Kap. 2); *Baron/Kerr*, Group Process 2003, 40; *Larson*, Synergy 2010, 63; *Drewes/Schultze/Schulz-Hardt* in: Frey/Bierhoff, Interaktion und Gruppe 2011, 221, 224.

[429] Die fünf Aufgabenarten werden üblicherweise zweistufig differenziert, näher die in der vorigen Fn. Genannten.

[430] *Laughlin* in: Fishbein, Progress 1980, 127; dazu *Larson*, Synergy 2010, 46 ff.

[431] Zu letzteren treffend *Larson*, Synergy 2010, 47: “Such tasks require an evaluative, behavioral or aesthetic judgment that establishes – as opposed to matches – what is correct.”

[432] Sie ermöglichen zwar aufschlussreiche Erkenntnisse über persönliche Konsistenz oder sozialen Einfluss (vgl. z.B. *Zarnoth/Sniezek*, J Exp Soc Psy 1997, 345, 353 f.), aber das liegt außerhalb der hier aufgeworfenen Fragen.

nehmensentscheidungen nichts hergebe, da diese nie „richtig“ oder „falsch“ seien.[433] Dem ist allerdings entgegenzuhalten, dass Unternehmensentscheidungen vielleicht nicht *demonstrierbar* richtig oder falsch sind, dass sie aber ebenso wenig „reine Ansichtssache“ sind. Wie bereits erwähnt, geht die Gruppenforschung von einem Aufgabenkontinuum aus, in dem die meisten Aufgaben zwischen den beiden Extremen angesiedelt sind. Viele Aufgaben sind so schwierig oder komplex, dass sie keine demonstrierbar richtige Lösung haben, wohl aber eine richtige oder zumindest bessere.[434] Solche Aufgaben sind es, die der Unternehmensalltag stellt, und sie lassen sich in kontrollierten Experimenten ohne Weiteres dadurch implementieren, dass dem Versuchsteilnehmer verborgen bleibt, ob es eine „richtige“ oder auch nur „bessere“ Lösung gibt.[435] Dass der Versuchs*leiter* gleichwohl weiß, was in der gegebenen Situation richtig wäre, ist gerade der Vorzug des kontrollierten Experiments. Während also alle nachfolgend dargestellten Studien in der Auswertung ihrer Ergebnisse von einer richtig/falsch-Dichotomie ausgehen, ist diese Dichotomie aus dem Versuchsaufbau normalerweise nicht erkennbar. Den Versuchsteilnehmern erscheinen ihre Entscheidungen also nicht weniger zweifels- oder wertungsbehaftet als kollegiale Entscheidungen im Unternehmensalltag.

Bezugspunkt. Schließlich ist konzeptionell klärungsbedürftig, mit welchem Einzelnen die Gruppenentscheidung verglichen wird.[436] Naheliegenderweise vergleicht die Gruppenforschung oft die Gruppenleistung direkt mit den Einzelleistungen der Gruppenmitglieder, testet also jedes Mitglied einmal im Einzel- und einmal im Gruppenversuch (*within subject design*). Da verschiedene Gruppenmitglieder allerdings unterschiedlich oft richtig entscheiden, haben sich zwei Vergleichsmaßstäbe etabliert. Diese müssen genau differenziert werden, da sie zwei verschiedene Interpretationen der

---

[433] *Michaelsen/Watson/Black*, J Appl Psy 1989, 834, 837: “in many organizational situations, the right answer is simply not known”; *Bainbridge*, Vand L Rev 2002, 1, 17; *Dorff*, Cardozo L Rev 2007, 2025, 2035: “the application of this research—which generally tests groups with questions that have an objectively correct answer—to corporate decisions—which generally do not—may turn out to be problematic”; allg. *Sniezek*, Org Behav Hum Dec Proc 1992, 124, 150: “objective evaluation is difficult and often impossible in many situations in which groups make decisions.”

[434] Vgl. *Laughlin*, Problem Solving 2011, 5: “Although the judgments […] had correct answers, […] correct answers in this research tradition are closer to the judgmental end of the intellective judgmental continuum”.

[435] Insoweit zutreffend *Bainbridge*, Vand L Rev 2002, 1, 17: “relevant experiments are those […] with respect to complex problems having a range of solutions.”

[436] Alternativ – und eingedenk der beliebten Metapher von der Gruppe als Informationsverarbeitungseinheit – lässt sich die Gruppenleistung an mathematischen Algorithmen messen (dazu noch bei Fn. 497 und 594); das versuchte namentlich *Davis*, Psy Rev 1973, 97 mit seinen Kombinationsmodellen (*social decision schemes*, SDS); ausf. *Kerr/Niedermeier/Kaplan*, Grp Proc Intergrp Rel 2000, 203.

Ursprungsfrage bedingen: Die erste Interpretation fragt, ob Gruppen besser entscheiden als ihr durchschnittliches bzw. typisches Mitglied. Ist das der Fall, lässt sich von schwacher Gruppensynergie (*weak synergy*) sprechen.[437] Die zweite Interpretation fragt, ob Gruppen besser entscheiden als ihr bestes Mitglied. Das wird unter dem Begriff der starken Gruppensynergie (*strong synergy*) behandelt,[438] die auch als Kreativitätsmehrwert (*creative plus*) oder, üblicher, als Gruppierungsbonus (*assembly effect bonus*) bekannt ist.[439] In disjunktiven Aufgaben ist ein solcher Gruppierungsbonus nur möglich, wenn die Gruppe mehr Alternativen berücksichtigt als jedes Gruppenmitglied, und wenn eine davon besser ist als die von jedem Gruppenmitglied einzeln bevorzugte. Mit anderen Worten: Ein Gruppierungsbonus in disjunktiven Aufgaben setzt voraus, dass kein Gruppenmitglied schon einzeln die beste Entscheidungsalternative identifiziert hatte.[440]

Die eingangs formulierte Hypothese, dass Gruppen „besser" entscheiden als Einzelne, lässt sich anhand dieser dritten Differenzierung in zwei Leitfragen übersetzen, bezogen jeweils auf (disjunktive) Entscheidungsaufgaben mit einer „besten" Alternative: Können Gruppen die unter ihren Mitgliedern verteilten Informationen zugunsten der besten Alternative erfolgreich zusammenzuführen? Und können sie die beste Alternative auch dann entdecken, wenn kein Gruppenmitglied sie von vornherein befürwortet? Beide Fragen sind nun nacheinander zu betrachten.

2. *Zusammenführung verteilter Informationen (hidden profile)*

Da in der Lebenswirklichkeit nie alle Mitglieder einer Gruppe vollständig informiert sind, hat die Gruppe potentiell Zugriff auf mindestens gleich viele, regelmäßig aber mehr Informationen als jedes einzelne ihrer Mitglieder.[441] Die potentielle Informationsbasis einer Gruppe ist damit theoretisch leicht zu ermitteln: sie entspricht der Summe aus allen geteilten (zentralen) und verteilten (dezentralen) Informationen. Schwieriger hingegen ist die empirische Frage, inwieweit es Gruppen auch tatsächlich gelingt, ihre verteilten Informationen zusammenzuführen, also ob „das prinzipiell vorhandene kognitive Problemlösungspotential auch tatsächlich für die Entscheidungsfundierung herangezogen wird."[442]

---

[437] *Larson*, Synergy 2010, 6.

[438] *Larson*, Synergy 2010, 7.

[439] *Larson*, Synergy 2010, 4 f. m.w.N.; mit letzterem meinten *Collins/Guetzkow*, Group Processes 1964, 58: "the group is able to achieve collectively something which could not have been achieved by any member working alone or by a combination of individual efforts".

[440] *Larson*, Synergy 2010, 67.

[441] *Sniezek*, Org Behav Hum Dec Proc 1992, 124, 131: "group members collectively have more information available than does any individual".

[442] *Grundei/von Werder*, AG 2005, 825, 829.

Betrachten wir ein stark vereinfachtes Beispiel: Wenn ein dreigliedriger Aufsichtsrat, der durch Mehrheitsbeschluss entscheidet (§ 108 AktG i.V.m. § 32 I 3 BGB analog), ein neues Vorstandsmitglied sucht und jedes der Aufsichtsratsmitglieder eine andere Information über einen bestimmten Bewerber hat – einer weiß um die guten Beziehungen, einer um den effektiven Führungsstil und der dritte um den tadellosen Ruf des Aspiranten – kennt kein Gruppenmitglied das vollständige Informationsprofil der Gruppe. Im Beispiel ist das unschädlich, weil alle drei Aufsichtsräte den Bewerber unabhängig voneinander befürworten. Was aber passiert, wenn ein Konkurrent auf den Plan tritt, dessen effektiver Führungsstil und tadelloser Ruf zwei von drei Aufsichtsräten bereits bekannt ist?[443] Dann befürwortet nur noch ein Aufsichtsratsmitglied den ersten Bewerber, denn aus Sicht der beiden anderen sprechen mehr Argumente (nämlich zwei statt nur eines) für den Konkurrenten. Obwohl die Aufsichtsratsmehrheit damit den Konkurrenten bevorzugt, sollte die Kollegialentscheidung am Ende dennoch lauten, den ersten Bewerber zu bestellen, da er auch über gute Beziehungen verfügt, die dem Konkurrenten fehlen.[444]

In dieser Situation – so die Annahme hinter dem Kollegialprinzip – hilft die kollegiale Gruppenberatung den drei Aufsichtsräten, ihre individuellen Informationen zusammenzutragen und die bestmöglich informierte Entscheidung zu treffen. Im Beispiel könnte etwa das Aufsichtsratsmitglied, das den ersten Bewerber bevorzugt, sein Argument vorbringen, woraufhin seine Kollegen feststellen würden, dass sie dieses Argument noch nicht kannten. Plötzlich spräche aus ihrer Sicht gleich viel für beide Bewerber. Nun muss nur noch eines der beiden Aufsichtsratsmitglieder, die zuvor den Konkurrenten bevorzugten, sein Wissen über den ersten Bewerber offenbaren – und schon wäre das verborgene Informationsprofil aufgedeckt und der Aufsichtsrat bevorzugte einstimmig den ersten Bewerber. Die Gruppe hätte damit das erreicht, was zuvor (bei Fn. 437) als schwache Synergie bezeichnet wurde: Ihre Entscheidung wäre besser als die Entscheidung des durchschnittlichen Gruppenmitglieds. Dieses Bild der Gruppendiskussion findet eine Stütze in der frühen sozialpsychologischen Theorie der argumentativen Überzeugung (*persuasive arguments theory, PAT*), nach der eine Gruppenberatung die Entdeckung verteilter Informationen begünstigt,

---

[443] Nun befürworten die meisten (aber nicht alle) Aufsichtsratsmitglieder den schwächeren Bewerber, daher ist das Informationsprofil zugunsten des besseren Bewerbers schwach verborgen (*weak hidden profile*), vgl. *Lu/Yuan/McLeod*, Pers Soc Psy Rev 2012, 54, 59; zu stark verborgenen unten nach Fn. 470.

[444] Anderes Beispiel (Akquisitionsentscheidung im Vorstand) bei *Brodbeck/Kerschreiter u.a.*, Acad Mgmt Rev 2007, 459, 461.

weil die geteilten Informationen als selbstverständlich hingenommen und nicht weiter diskutiert werden.[445]

Ebenso gut könnte es im Beispiel aber passieren, dass die Gruppenberatung mit den Argumenten eröffnet wird, die einer der beiden Befürworter des Konkurrenten parat hat. Dann würden beide Befürworter feststellen, dass sie identische Informationen haben – und sich in ihrem Vertrauen auf diese Informationen bestärkt sehen. Das dritte Mitglied ließe sich womöglich umstimmen, und niemand sähe mehr Anlass, noch über den ersten Bewerber zu sprechen. Diese Vorstellung der typischen Gruppenberatung hat sich nach zahlreichen Studien an verborgenen Informationsprofilen (*hidden profiles*) seit 1985 durchgesetzt.[446] In diesen Studien wurden Versuchsgruppen mit Situationen wie der hier beschriebenen konfrontiert, in denen unter den Gruppenmitgliedern Informationen verteilt sind, die eine bessere Entscheidung ermöglichen als die geteilten Informationen der Gruppenmitglieder. Viele Versuchsgruppen lassen sich durch ihre geteilten Informationen einnehmen, durchmustern ihre verteilten Informationen nur noch einseitig (*biased information sampling*) und entdecken das verborgene Informationsprofil nicht;[447] zumeist bestärken sie sich stattdessen gegenseitig in der Illusion, bestmöglich informiert gewesen zu sein.[448]

Unlängst wurden 65 solcher Versuche mit verborgenen Informationsprofilen (durchgeführt zwischen 1985 und 2010) in einer Metastudie ausgewertet.[449] Die Metastudie kommt zu dem Ergebnis, dass in Fällen wie dem oben gebildeten Beispiel die Gruppenberatung in der Tat durchweg mehr geteilte (*common*) als verteilte (*unique*) Informationen zutage fördert.[450] Gruppen finden die beste Handlungsoption dreimal so selten, wenn diese nur anhand verteilter Informationen zu erkennen ist, als wenn die erforderlichen Informationen allen Gruppenmitgliedern von vornherein zur Verfügung stehen.[451] Die Metastudie untersucht eine Reihe weiterer Einflussfaktoren und kommt zu der Schlussfolgerung, dass die richtige Entscheidung am seltensten gewählt wird, wenn die dafür erforderlichen Informationen

---

[445] *Burnstein/Vinokur*, J Exp Soc Psy 1977, 315, 316 f.; vgl. schon *Burnstein/Vinokur/Trope*, J Exp Soc Psy 1973, 236.

[446] Erstmals *Stasser/Titus*, J Pers Soc Psy 1985, 1467; ideengeschichtlicher Abriss *Stasser/Titus*, Psy Inq 2003, 304.

[447] *Gigone/Hastie*, J Pers Soc Psy 1993, 959 prägten dafür das Schlagwort des *common knowledge effect.*

[448] *Baron/Kerr*, Group Process 2003, 105.

[449] *Lu/Yuan/McLeod*, Pers Soc Psy Rev 2012, 54, die auch zwei frühere Metastudien (*Mesmer-Magnus/DeChurch*, J Appl Psy 2009, 535; *Reimer/Reimer/Czienskowski*, Comm Mono 2010, 121) replizieren; daher übergehe ich beide.

[450] *Lu/Yuan/McLeod*, Pers Soc Psy Rev 2012, 54, 65 f., 69: “on average two standard deviations more of the common information than of the unique information is pooled”.

[451] *Lu/Yuan/McLeod*, Pers Soc Psy Rev 2012, 54, 66 Tabelle 7: 1 / 2,62 = 38,17 %.

nicht nur verteilt, sondern auch noch schwer intersubjektiv vermittelbar sind, etwa weil die beste Entscheidung nicht leicht als solche zu erkennen ist.[452] Eines der in der Metastudie ausgewerteten Experimente belegt zudem den gravierenden Effekt, den es haben kann, wenn die in der Gruppe verteilten Informationen nicht unabhängig voneinander sind, sondern aufeinander aufbauen (*interdependent information*).[453] Da verteilte Informationen auch umso seltener erwähnt werden, je größer die Gruppe oder die Gesamtmenge an Informationen ist,[454] nimmt das Problem wohl ganz generell mit der Komplexität der Entscheidung zu.

Zweifel an der *praktischen* Relevanz verborgener Informationsprofile könnten sich indessen aus der Teamforschung zur verteilten Informationsverarbeitung (*shared cognition*) ergeben. Anfangs wurden dafür ganz verschiedene Konzeptionen und Begriffe entwickelt,[455] derzeit am weitesten durchgesetzt hat sich die Theorie des mentalen Teammodells (*team mental model, TMM*).[456] Sie besagt, dass jedes Teammitglied mit der Zeit Erwartungen sowohl an die Aufgaben des Teams als auch an das Team selbst bildet, die es ihm zunehmend erleichtern, sich in andere Teammitglieder hineinzuversetzen und mit ihnen unausgesprochen zu koordinieren.[457] Diese wechselseitige Koordination erzeuge auf Ebene des Teams emergente Repräsentationen der Aufgaben und Zusammensetzung des Teams, die die Entscheidungsqualität maßgeblich beeinflussen.[458] Der für den vorliegen-

---

[452] *Lu/Yuan/McLeod*, Pers Soc Psy Rev 2012, 54, 70: "Taken together, the results of these three meta-analyses [vgl. Fn. 450] indicate that hidden profile tasks without a clear preferred solution are most detrimental".

[453] *Fraidin*, Org Behav Hum Dec Proc 2004, 102, 108: Dann finden noch einmal 35 % weniger Gruppen die Lösung.

[454] *Lu/Yuan/McLeod*, Pers Soc Psy Rev 2012, 54, 70: "every additional group member was found to increase the gap between mentions of common and unique information by 0.32 standard deviations […] every additional piece of information increased [… it] by 0.003 standard deviations"; zu den Parallelen zur sog. Gruppenpolarisierung vgl. below IV.2..

[455] *Klimoski/Mohammed*, J Mgmt 1994, 403: "Terms such as shared mental models, common cause maps, shared frames, teamwork schemas, transactional memory, and sociocognition are being offered by investigators" m.w.N. in Tabelle 1, und weiter 407: "Even terminology like 'teamthink' is being bandied about in the popular press"; viele weitere Bsp. bei *DeChurch/Mesmer-Magnus*, J Appl Psy 2010, 32, 39 Tabelle 1.

[456] Krit. noch *Klimoski/Mohammed*, J Mgmt 1994, 403, 413: "the concept of shared mental models is often merely invoked post hoc".

[457] *Klimoski/Mohammed*, J Mgmt 1994, 403, 412: "Teams who share mental models are expected to have common expectations of the task and team, allowing them to predict the behavior and resource needs of team members more accurately"; *Kim*, Org Behav Hum Dec Proc 1997, 165, 166; *Kozlowski/Ilgen*, Psy Sci Publ Interest 2006, 77, 81: "Process begets structure, which in turn guides process."

[458] Näher zum TMM *Klimoski/Mohammed*, J Mgmt 1994, 403; *Mohammed/Ferzandi/Hamilton*, J Mgmt 2010, 876; *DeChurch/Mesmer-Magnus*, Grp Dyn 2010, 1 (Metastudie über 23 Primärstudien).

den Zusammenhang wichtigste Teil eines solchen mentalen Modells wird als transaktives Gedächtnis (*transactive memory*) bezeichnet und meint das in einer Gruppe „geteilte Bewusstsein darüber, wer in der Gruppe was weiß“.[459] Wenn beispielsweise ein Vereinsvorstand über ein Projekt entscheiden muss, für das je zehn Informationen aus den Bereichen Finanzen, Öffentlichkeitsarbeit und Mitgliederverwaltung relevant sind, muss jeder der drei Vorstände für Finanzen, Öffentlichkeitsarbeit und Mitgliederverwaltung nur zwölf Informationen vorhalten: zehn aus seinem Ressort sowie die beiden Tatsachen, dass er zehn weitere Informationen jeweils vom zweiten und dritten Vorstandskollegen erhält. Der Gesamtvorstand hat dann Zugriff auf dreißig Informationen, von denen kein Mitglied auch nur die Hälfte präsent halten muss; stattdessen kann sich jeder auf sein Ressort konzentrieren, mit dessen Informationen er zweifellos am besten umgehen kann. Soweit Gruppen in der Praxis auch ohne explizite Aufgabenverteilung solche transaktiven Gedächtnisse ausbilden – was empirisch schon in mehreren Fällen belegt wurde[460] – könnten verborgene Informationsprofile der oben beschriebenen Art von vornherein ausbleiben. Implizite oder explizite Aufgabenteilung sollte dazu führen, dass Informationen nicht wahllos, sondern nach den Kompetenzen der Gruppenmitglieder verteilt und dadurch für die anderen Gruppenmitglieder leichter auffindbar sind. Ein Experiment, das ein verborgenes Informationsprofil mit just dieser Art der kognitiven Arbeitsteilung kombinierte,[461] kam zu dem Ergebnis, dass Gruppen, deren Mitglieder auf bestimmte Informationen spezialisiert waren, fast doppelt so häufig zur richtigen Entscheidung kamen wie Gruppen ohne solche Spezialisierung – und fast so oft wie Gruppen, deren Mitglieder alle relevanten Informationen von vornherein zur Verfügung hatten.[462]

Ob transaktive Gedächtnisse in der Praxis tatsächlich der Entstehung von verborgenen Informationsprofilen vorbeugen können, erscheint allerdings aus mehreren Gründen zweifelhaft. Zunächst muss den Gruppenmitgliedern hinreichend genau bekannt sein, welcher ihrer Kollegen eine Präferenz oder Kompetenz für welche Art von Informationen hat.[463] Das mag im eben gebildeten Beispiel des Vereinsvorstands mit ressortbezogenen In-

---

[459] *Peltokorpi*, Rev Gen Psy 2008, 378: “shared awareness of who knows what within the group”; dazu auch *Ilgen/Hollenbeck u.a.*, Ann Rev Psy 2005, 517, 526; vgl. schon den Titel von *Stasser/Stewart/Wittenbaum*, J Exp Soc Psy 1995, 244.

[460] Schilderung einiger Studien bei *Larson*, Synergy 2010, 244 ff.; Forschungsbericht bei *Peltokorpi*, Rev Gen Psy 2008, 378; vgl. auch *Stasser/Stewart/Wittenbaum*, J Exp Soc Psy 1995, 244, 246: “groups that have a history of working together may develop a division of responsibility for obtaining, processing and communicating certain types of information”.

[461] *Stasser/Stewart/Wittenbaum*, J Exp Soc Psy 1995, 244.

[462] *Stasser/Stewart/Wittenbaum*, J Exp Soc Psy 1995, 244, 251 f.: 61 % bzw. 35 % bzw. 70 %; für den ersten Vergleich $\chi^2(1, N = 73) = 4{,}96$, $p < 0{,}03$, für den zweiten $\chi^2(1, N = 59) = 0{,}45$, n.s.

formationen (oder in einem nach Sparten gegliederten AG-Vorstand) weniger problematisch sein, wird aber bei vielschichtigen Informationen über komplexe Sachverhalte deutlich schwieriger. Solche ressortübergreifenden Informationen in einem transaktiven Gedächtnis abrufbereit zu halten – also zu verhindern, dass jeder sich auf seine Kollegen verlässt und letztlich niemand die Information behält – ist eine anspruchsvolle Koordinationsleistung. Bedenkt man zusätzlich, dass Menschen dazu neigen, den Kenntnisstand ihrer Mitmenschen am eigenen Wissen zu messen und dadurch zu überschätzen (*curse of knowledge*), könnte sich das transaktive Gedächtnis mit zunehmender Vertrautheit der Gruppe sogar verschlechtern.[464]

In einem Experiment zur Überprüfung dieser Hypothese schnitten Gruppen ausgerechnet dann am schlechtesten ab, wenn sie sowohl mit ihrer Aufgabenstellung als auch mit ihren Gruppenkollegen bereits vertraut waren.[465] Das schließt zwar nicht aus, dass eine noch längere oder intensivere Zusammenarbeit zur Herausbildung von erfolgreichen transaktiven Gedächtnissen führt.[466] Daran lässt aber eine weitere Studie zweifeln, die 45-Tages-Vorhersagen des Ölpreises durch eine eingespielte Gruppe von sechs Entscheidungsträgern in einem internationalen Erdölunternehmen untersucht und festgestellt hatte, dass die Gruppenvorhersage schlechter ausfiel als der Durchschnitt der sechs isoliert erhobenen Einzelvorhersagen.[467] Selbst wenn aber Gruppen in der Praxis von ihrem transaktiven Gedächtnis profitieren, riskieren sie dafür umgekehrt eine erhöhte Selbstüberschätzung.[468]

3. Erweiterung des Alternativenspektrums (assembly effect bonus)

Der vorangegangene Abschnitt widmete sich der Frage, inwieweit Gruppen verteilte Informationen zusammenführen können. Dabei wurde die potentielle Informationsbasis der Gruppe als Summe der geteilten und verteilten

---

[463] *Stasser/Stewart/Wittenbaum*, J Exp Soc Psy 1995, 244, 262; *Peltokorpi*, Rev Gen Psy 2008, 378, 379.

[464] *Kim*, Org Behav Hum Dec Proc 1997, 165, 167: "those who have gained experience with their partners and/or task would not have their own lack of familiarity as a reminder that others may not know what they know and they may, therefore, form less conservative assumptions about their partners' knowledge."

[465] *Kim*, Org Behav Hum Dec Proc 1997, 165, 172, in einer 2x2-Variation – also mit drei weiteren Versuchsbedingungen, deren Teilnehmer nur mit der Aufgabe, nur mit ihren Kollegen oder mit keinem von beiden vertraut waren.

[466] So auch *Kim*, Org Behav Hum Dec Proc 1997, 165, 174.

[467] *Brocato/Kumar/Smith*, Manage Dec Econ 1989, 13, 22: "The evidence, using the rational expectations tests as criteria, shows group-produced forecasts inferior to the mathematical average of individually produced predictions."

[468] *Peltokorpi*, Rev Gen Psy 2008, 378, 379; zur Selbstüberschätzung in Gruppen noch ausf. unten V.

Informationen bezeichnet. Ausgeblendet blieb dabei zunächst, dass die Gruppendiskussion ja auch ganz neue Erkenntnisse erzeugen, also das Spektrum der Entscheidungsalternativen erweitern kann. Soweit das passiert, kann eine Gruppe ihr bestes Mitglied übertreffen und dadurch starke Synergie verwirklichen – den sog. Gruppierungsbonus (*assembly effect bonus*, vgl. oben bei Fn. 439). Im oben gebildeten Beispiel des Aufsichtsrats war dieser Gruppierungsbonus gar nicht möglich, denn eines der Aufsichtsratsmitglieder befürwortete bereits vor der Beratung den besseren Vorstandsbewerber. Die Gruppe konnte also keine neuen Entscheidungsalternativen mehr generieren und – was gleichbedeutend ist – ihr bestes Mitglied nicht übertreffen.[469] Bestenfalls hätte die Gruppe ihr typisches Mitglied (nämlich die beiden Aufsichtsratsmitglieder, die zunächst den schlechteren Vorstandsbewerber bevorzugten) übertreffen und schwache Synergie realisieren können. Das ändert sich, wenn wir das Beispiel geringfügig abwandeln: Vor der Beratung kannten nicht nur zwei, sondern alle drei Aufsichtsräte die beiden Vorzüge des konkurrierenden Bewerbers. In diesem Fall hätte niemand den ersten Bewerber befürwortet, er wäre also gar nicht im Alternativenspektrum der Gruppe aufgetaucht. Inwieweit kann die Gruppenberatung diesen Bewerber trotzdem „ins Spiel bringen“ und feststellen, dass die überwiegenden Argumente für ihn sprechen?

Eine solche Weitung der Perspektive wäre natürlich der größtmögliche Gewinn, den das Kollegialprinzip erbringen kann – allerdings auch der unwahrscheinlichste.[470] Denn während der Vorstandsbewerber im ersten Beispiel zumindest noch einen Befürworter im Aufsichtsrat hatte, scheint er in der hier gebildeten Variante keinem Aufsichtsratsmitglied ernsthaft vorzugswürdig und steht deshalb womöglich gar nicht „zur Debatte“. Ob es Gruppen unter diesen Umständen dennoch gelingt, die verborgene Entscheidungsalternative zu entdecken und zu verfolgen, war eine weitere Frage der oben (bei Fn. 449) zitierten Metastudie. Sie differenzierte die untersuchten Studien danach, ob sie verborgene Informationsprofile mit schwacher Ausprägung verwendeten (*weak hidden profile*), wie im ursprünglichen Beispiel, oder mit starker Ausprägung (*strong hidden profile*), wie im abgewandelten Beispiel. Der Unterschied ist erheblich: Gruppen, deren Informationsprofil stark verborgen war, wählten sechseinhalb Mal seltener die beste Alternative als Gruppen mit schwach verborgenem Infor-

---

[469] Dass das „beste“ Mitglied sich hier durch *Unkenntnis* der geteilten Informationen qualifiziert, ist der Einfachheit des Beispiels geschuldet; in komplexeren Situationen könnte das beste Mitglied etwa mehr Gegenargumente wissen.

[470] So meint *Ueberwasser*, Kollegialprinzip 1989, 46 m.w.N.: „Zu idealisierend ist auch die Vorstellung, […] «Durch das Zusammenstrahlen [mehrerer] Subjektivitäten» bilde sich dann eine «Gesamtmeinung» heraus, die sehr häufig keiner für sich allein gefunden hätte.“

mationsprofil, also in nur noch knapp 6 % derjenigen Fälle, in denen eine Gruppe mit voll informierten Mitgliedern die beste Alternative wählt.[471]

Wiewohl 6 % keine ganz zu vernachlässigende Größe ist, bieten Studien zur Informationsaggregation also eher geringen Anhalt für eine Perspektivenweitung durch Gruppenarbeit. Insoweit harmonieren die referierten Studien mit anderen Bereichen der *Gruppen*forschung.[472] Dennoch finden sich in der *Team*forschung immer wieder Zweifel an diesem Befund. Oft beobachten Managementpraktiker und Teamforscher, dass Arbeitsgruppen und andere Teams im Alltag durchaus bessere Entscheidungen treffen als Einzelne, und schließen daraus, dass Laborsituationen schlicht zu lebensfremd seien, um verwertbare Erkenntnisse über das Gruppenverhalten zu liefern. Eine solche Studie, die besondere Aufmerksamkeit erfahren hat, wurde 1989 mit der erklärten Absicht veröffentlicht, endlich einen „realistischen Vergleich der Entscheidungsfindung bei Individuen und Konsensgruppen“ beizutragen.[473] Die Studie wurde später von zwei renommierten Gruppenforschern kritisiert, die in einer Replik aufzeigten, dass die vermeintlich spektakulären Ergebnisse der Studie schlicht auf mangelhafter Datenauswertung beruhten.[474] Der Replik folgt eine Duplik der Autoren[475] und schließlich gar noch eine Triplik der Kritiker[476] sowie später eine weitere kritische Stellungnahme anderer Autoren.[477] Trotz dieser heftigen Kontroverse wurde die Ursprungsstudie im Gesellschaftsrecht für bare Münze genommen und als „interessantes Experiment, dessen Ergebnisse nahelegen, dass Gruppenentscheidungen Synergie-Effekte haben“ rezipiert – ohne die späteren Einwände auch nur zu erwähnen.[478] Es bedarf also einer näheren Auseinandersetzung mit dieser Studie, um ihre Ergebnisse einordnen und informiert bewerten zu können.

Fragestellung. Die Ursprungsstudie führte den Mangel an experimentellen Belegen für einen Gruppierungsbonus darauf zurück, dass die experimentelle Gruppenforschung in drei Dimensionen unrealistische Situationen teste. Zum einen würden Gruppen im Labor *ad hoc* und „nur für die Dauer

---

[471] *Lu/Yuan/McLeod*, Pers Soc Psy Rev 2012, 54, 66 Tabelle 7: 17,09/2,62 = 6,523 und 1 / 17,09 = 0,0585.

[472] Ausf. *Larson*, Synergy 2010, 7: “In contrast to weak synergy, there is relatively sparse evidence for the existence of strong synergy, especially on tasks that require mental rather than physical effort.”; ebenso *Hill*, Psy Bull 1982, 517, 535.

[473] So der Titel von *Michaelsen/Watson/Black*, J Appl Psy 1989, 834.

[474] *Tindale/Larson*, J Appl Psy 1992, 102: “analyzed their data inappropriately”; zust. *Gigone/Hastie*, Psy Bull 1997, 149, 153: “mistakenly […] equivocal method of analysis”.

[475] *Michaelsen/Watson u.a.*, J Appl Psy 1992, 106.

[476] *Tindale/Larson*, J Appl Psy 1992, 109; vgl. auch *Larson*, Synergy 2010, 33–38.

[477] *Stasson/Bradshaw*, Small Grp Res 1995, 296.

[478] *Bainbridge*, Vand L Rev 2002, 1, 24 f.; ähnl. *Pollman*, UC Davis Bus L J 2009, 137, 150 Fn. 48; *Charness/Sutter*, J Econ Persp 2012, 157, 172.

der Datenerhebung – normalerweise weniger als 1 Stunde“ zusammengestellt, zum zweiten seien die Entscheidungsaufgaben nicht „im einschlägigen Zusammenhang“ (gemeint wohl: unternehmerische Entscheidungen) angesiedelt und den Teilnehmern nicht hinreichend vertraut, und drittens fehle meist der finanzielle Anreiz, das Problem richtig zu lösen und falsche Lösungen zu vermeiden – all diesen Mängeln wollte die Studie abhelfen, indem sie „Daten von erfahrenen Gruppen sammelte, die im einschlägigen Zusammenhang folgenreiche Entscheidungen“ getroffen hatten.[479]

Untersuchungsplan. Die Autoren erhoben Daten von 1.334 Individuen, die innerhalb von fünf Jahren in 25 Durchgängen eines BWL-Kurses insgesamt 222 Lerngruppen zugeteilt worden waren. 85 % der Kursteilnehmer stammten aus der Studentenpopulation der betreffenden Universität, 12 % aus dem weltweiten Austauschprogramm, 2 % aus einem betrieblichen Fortbildungsprogramm. Die Lerngruppen wurden zu Beginn des Kurses gezielt so eingeteilt, dass alle Gruppen über möglichst gleiche Kompetenzen verfügten.[480] Alle Gruppen hatten zwischen drei und acht Mitgliedern; im Durchschnitt bestand eine Gruppe aus sechs Mitgliedern (Ø 5,97 ± 0,84).[481] Die Gruppen lernten gemeinsam für mindestens 27 Zeitstunden während des Kurses, und nach Schätzung der Autoren noch einmal so lange außerhalb des Kurses. Zudem verbrachten sie insgesamt etwa fünf Stunden damit, gemeinsam sechs Prüfungen zu bearbeiten, die zu Beginn jedes Themenblocks gestellt wurden, um das Verständnis der vorab aufgegebenen Heimlektüre zu testen. Dabei musste jeder Teilnehmer den Prüfungsbogen (12-18 Auswahlfragen) zunächst selbst ausfüllen und abgeben, und sofort danach noch einmal denselben Prüfungsbogen in der Gruppe bearbeiten. Etwa 40 % der Auswahlfragen testeten die Merkleistung, 40 % das Anwendungsverständnis, 20 % das synthetisch-analytische Verständnis.[482] Die Ergebnisse der sechs Prüfungen zählten nach einem von den Studenten selbst festgelegten Gewichtungsschema – durchschnittlich trugen die Einzelprüfungen 10 % und die Gruppenprüfungen 15 % zur Gesamtnote des Kurses bei.

Auswertung. Die Autoren der Studie verglichen für jeden Kursteilnehmer die Gesamtpunktzahl der sechs Einzelprüfungen mit der Gesamtpunktzahl der sechs Gruppenprüfungen. Da die sechs Prüfungen nicht einheitlich

---

[479] *Michaelsen/Watson/Black*, J Appl Psy 1989, 834, 835.

[480] *Michaelsen/Watson/Black*, J Appl Psy 1989, 834, 835: Der Kursleiter ließ die Teilnehmer nach einem bestimmten Kompetenzkriterium (etwa Berufserfahrung oder fachlicher Hintergrund) aufstehen, teilte ihnen zufällige Gruppennummern zu und rief anschließend ein anderes Kompetenzkriterium auf – solange, bis alle Teilnehmer zugeteilt waren.

[481] Hier und im Folgenden steht Ø für den Mittelwert und ± für die Standardabweichung (zu beiden vgl. § 2 A.IV.1.).

[482] Beispiele bei *Michaelsen/Watson/Black*, J Appl Psy 1989, 834, 839 (Appendix).

lang waren, konnten Prüflinge insgesamt zwischen 84 und 101 Punkten erzielen. Die Autoren beobachteten, dass der *durchschnittliche* Kursteilnehmer in den Einzelprüfungen insgesamt 74,2 Punkte (± 5,9) erzielte, die jeweils *gruppenbesten* Teilnehmer Ø 82,6 Punkte (± 6,1) und die Gruppen Ø 89,9 Punkte (± 5,2). Ein statistischer Vergleich zwischen diesen Werten ergab,[483] dass die Gruppe signifikant besser abschnitt als ihre einzelnen Mitglieder, und zwar 21,2 % besser als ihr durchschnittliches und 8,8 % besser als ihr bestes Mitglied. Nur in drei von 222 Gruppen war das beste Mitglied seiner Gruppe überlegen. Die Gruppen*größe* hatte einen widersprüchlichen Einfluss, den die Autoren nicht weiter besprachen, sondern als irrelevant verwarfen.[484]

Deutung. Die Autoren interpretieren ihre Ergebnisse als „unzweideutigen Beleg, dass in einer Situation, die einer typischen Arbeitsumgebung in vielerlei Hinsicht ähnelt, die überwältigende Mehrzahl der Gruppen bessere Entscheidungen treffen kann als ihr kompetentestes Mitglied."[485] Das illustriere deutlich, dass Gruppenforschung unter Bedingungen stattfinden müsse, die so realistisch wie möglich seien. Dass diese Bedingungen andere Ergebnisse erbracht hätten als die frühere Gruppenforschung, deuten die Autoren als Resultat verschiedener Faktoren: Erstens könne die deutlich höhere Gruppenvertrautheit dazu geführt haben, dass die intellektuellen Ressourcen der Mitglieder besser genutzt wurden. Anders als in einer auf Dauer angelegten Organisation sei die Gruppenarbeit mit Abschluss des Kurses zwar abrupt beendet worden, was eher einer Projektgruppe ("task force") ähnele, aber das lasse vermuten, dass echte Organisationen zu einem noch größeren Gewinn an Entscheidungsqualität führen.[486] Zweitens seien die Ergebnisse durch den in der Benotung liegenden Anreiz verursacht, die richtige Lösung zu finden. Die Autoren bezweifeln zwar, dass dieser Anreiz den Gehalts- und Beförderungsaussichten am Arbeitsplatz vergleichbar sei, schlagen aber vor, solche Anreizsysteme auch mit Blick auf die Ergebnisse des Experiments auszugestalten. Drittens diskutieren die Autoren den „wahrscheinlich gewichtigsten Grund für mögliche Bedenken gegen die Verallgemeinerung der Ergebnisse",[487] die Aufgabenstellungen selbst. Diese seien der Informationsverarbeitung in Organisationen zwar in wichtigen Aspekten recht ähnlich, wiesen aber eine deutlich geringere Komplexität und Bandbreite auf. Zudem sei es in der Realität oft gar

---

[483] $\chi^2(2, N = 222) = 444$ für das durchschnittliche, $\chi^2(2, N = 222) = 403$ für das beste Mitglied, jeweils $p < 0{,}01$.

[484] *Michaelsen/Watson/Black*, J Appl Psy 1989, 834, 838: "no significant correlation between group size and an assembly bonus effect".

[485] *Michaelsen/Watson/Black*, J Appl Psy 1989, 834, 836.

[486] *Michaelsen/Watson/Black*, J Appl Psy 1989, 834, 837.

[487] *Michaelsen/Watson/Black*, J Appl Psy 1989, 834, 837.

nicht möglich, Entscheidungen als falsch oder richtig zu bewerten, weshalb die Ergebnisse nur beschränkt übertragbar seien. Sie sprächen aber doch dafür, im Unternehmenskontext auf Gremien zurückzugreifen, um möglichst viele Entscheidungsaspekte zu Tage zu fördern. Zuletzt weisen die Autoren darauf hin, dass die in ihrer Studie ausgewerteten BWL-Kurse gerade das Verhalten in Organisationen zum Gegenstand hatten, weshalb die Teilnehmer ein größeres Bewusstsein für die Erfolgsfaktoren von Gruppenentscheidungen gehabt haben dürften als der normale Entscheidungsträger; andererseits sei ein ähnliches Bewusstsein für Gruppenkonzepte in realen Organisationen aufgrund von Trainings und Programmen zur Teamentwicklung angelegt.[488]

Diskussion. Die methodische Erörterung der Studie wird zunächst dadurch erschwert, dass in der gesamten Studie nicht klar wird, ob die Autoren absolute Punktzahlen in der Prüfung oder den prozentualen Anteil der richtig beantworteten Aufgaben verwenden und berichten. Da die Prüfungen unterschiedlich lang und die Gruppen unterschiedlich groß waren, sind nur Prozentanteile sinnvoll interpretierbar. Andererseits heißt es im Text unmissverständlich, die „Daten bestehen aus kumulierten Punktzahlen […] von 84 bis 101 insgesamt“,[489] und die Ergebnisse werden als „Durchschnitt der Punktzahlen“ und ohne %-Zeichen berichtet.[490] Dementsprechend habe ich oben die Ergebnisse als absolute Werte wiedergegeben, was aber die Vergleichbarkeit der berichteten Daten einschränkt und die berichteten Auswertungen jeder Nachvollziehbarkeit beraubt. Immerhin ergab eine Replikationsstudie mit geringer Stichprobengröße (8 Gruppen), dass durchschnittliche Individuen 72,2 %, die besten Individuen 85,1 % und Gruppen 90,1 % der Fragen richtig beantworteten; da diese Zahlen nominell nahe an den von den Autoren berichteten Werten liegen, gingen die Urheber der Replikationsstudie ohne Weiteres davon aus, die Autoren hätten Prozentwerte berichtet.[491]

Schließt man sich dieser Deutung einmal an, verbleiben zwei gewichtige Kritikpunkte: Zum einen ist unklar, ob die höhere Gesamtpunktzahl der Gruppen nicht allein durch einen Übungseffekt statt durch einen Gruppie-

---

[488] *Michaelsen/Watson/Black*, J Appl Psy 1989, 834, 837 f.: “an awareness of group process issues may have been a significant aid to the groups in our study […] however, it would seem that similar conditions could be created in ongoing organizations through appropriately focused training or organizational development programs.”

[489] *Michaelsen/Watson/Black*, J Appl Psy 1989, 834, 835: “Data consisted of the cumulative scores obtained from a series of six individual and group tests. Each of these tests contained 12 to 18 multiple-choice and true/false questions, for a total of 84 to 101 items in all.”

[490] *Michaelsen/Watson/Black*, J Appl Psy 1989, 834, 836: “mean group score […] mean average individual score […] group scores and best individual scores”.

[491] *Tindale/Larson*, J Appl Psy 1992, 102, 103; *Larson*, Synergy 2010, 34; ähnl. auch *Stasson/Bradshaw*, Small Grp Res 1995, 296, 297.

rungsbonus zu erklären war. Zum anderen kann der Vergleich von Gesamtpunktzahlen ohnehin nie einen Gruppierungsbonus (also starke Synergie) belegen, sondern allenfalls schwache Synergie. Der erste Einwand ist naheliegend: Da in der Ursprungsstudie jeder Kursteilnehmer von der Einzelprüfung in die Gruppenprüfung wechselte – dort also nicht nur *als Gruppenmitglied*, sondern auch *zum zweiten Mal* entschied – könnte die höhere Gruppenpunktzahl schlicht an der wiederholten Übung liegen, nicht dagegen an der Gruppenarbeit als solcher. Diesem Identifikationsproblem hätte eine Kontrollbedingung vorbeugen können, in der Teilnehmer dieselben Aufgaben zweimal hintereinander allein bearbeiten. Eine spätere Replikationsstudie hat genau das versucht, und konnte statistisch nicht ausschließen, dass Gruppen allein wegen des Übungseffekts besser abschnitten.[492] Die Kursteilnehmer hätten also beim zweiten Versuch wahrscheinlich auch ohne Gruppenarbeit besser abgeschnitten.

Selbst wenn aber – was nur noch reine Spekulation sein kann – die Gruppen nicht allein wegen des Übungseffekts höhere Gesamtpunktzahlen erzielt hätten, belegt das keineswegs einen Gruppierungsbonus. Dieser würde vorliegend ja voraussetzen, dass die Gruppe Aufgaben richtig beantworten konnte, die keines ihrer Mitglieder zuvor richtig beantwortet hatte. Mit anderen Worten: Für die Schlussfolgerung (oben bei Fn. 485), dass die „Mehrzahl der Gruppen bessere Entscheidungen treffen kann als ihr kompetentestes Mitglied“, müssten schon die Entscheidungen der Gruppe mit denen des kompetentesten Mitglieds verglichen werden. Gesamtpunktzahlen dagegen sind keine Entscheidungen der Versuchsteilnehmer, sondern eine Entscheidung des Versuchsleiters, die Entscheidungen der Versuchsteilnehmer nach einem bestimmten Verfahren zu aggregieren.[493] Ob nämlich eine, zwei oder 101 Entscheidungen von Individuum und Gruppe für den Vergleich aggregiert werden, ist eine willkürliche Setzung ohne Bezug zu den real getroffenen Entscheidungen. Anders formuliert: Der Vergleich von Gesamtpunktzahlen erlaubt lediglich eine Aussage darüber, wie die Gruppe im Vergleich zu ihrem *durchschnittlich* besten Mitglied abschneidet (d.h. schwache Synergie),[494] kann aber keine Erweiterung des Alternativenspektrums (d.h. Gruppierungsbonus) belegen.

---

[492] *Stasson/Bradshaw*, Small Grp Res 1995, 296, 299, 304: “groups did perform significantly better than the highest scoring individual, but […] not significantly above the […] improvement rate obtained through re-testing. Thus the practice effect hypothesis could not be rejected.”

[493] So auch *Larson*, Synergy 2010, 35: “The students, whether working individually or as a group, affect only *what* is aggregated – exam item answers – not *how* those answers are aggregated. […] it is really the instructor, not the students, who creates the total exam scores.”

[494] *Larson*, Synergy 2010, 37: das laut der Ursprungsstudie „beste“ Mitglied sei “really only ‘best on average’”.

Statt der Gesamtpunktzahl hätten also die Einzelantworten ausgewertet werden müssen.[495] Da diese nach Aussage der Autoren überhaupt nicht dokumentiert wurden, kann die Ursprungsstudie über einen Gruppierungsbonus allenfalls spekulieren.[496] Immerhin lässt sich mittels Computersimulation ausrechnen, dass sechs Individuen, die durchschnittlich – wie in der Ursprungsstudie berichtet – 74,2 % der Fragen richtig beantworten, in der Gruppe 99,5 % der Fragen richtig beantworten sollten, wenn sich die Gruppe von zwei richtig Antwortenden überzeugen lässt.[497] Dass die von der Ursprungsstudie beobachtete Gruppenerfolgsquote von 89,9 also auf Gruppierungsboni beruhe, lässt sich durchaus bezweifeln.[498] So ergab auch die erste spätere Replikationsstudie, dass Gruppen, in denen kein Mitglied anfangs die richtige Lösung wusste, nie zu dieser Lösung gelangten – und selbst wenn ein Mitglied die richtige Lösung wusste, noch in zwei Drittel der Fälle zur falschen Lösung kamen, solange nur drei oder vier andere Gruppenmitglieder die Aufgabe falsch gelöst hatten.[499] Dennoch waren die Gruppen auch in jener Studie *insgesamt* fast fünf Prozentpunkte besser als ihr durchschnittlich bestes Mitglied.

Die Autoren der Ursprungsstudie verteidigten ihre Auswertungsmethode vehement damit, dass in der Realität selten einzelne Aufgaben zur Bewertung anstünden, sondern meist Probleme mit zahlreichen Facetten auf mehreren Wissensebenen.[500] Diese Argumentation verkennt völlig die von der Kritik aufgeworfene Differenzierung zwischen schwacher und starker Synergie und illustriert, dass die gesamte Debatte wohl auf einem Missverständnis des Begriffs *assembly effect bonus* beruht.[501] Während die Gruppenforschung ihn als Fachbegriff für die sehr seltenen Fälle starker Synergie verwendet, bezogen die Autoren der Ursprungsstudie ihn offenbar auf

---

[495] *Tindale/Larson*, J Appl Psy 1992, 102, 105; zust. *Gigone/Hastie*, Psy Bull 1997, 149, 153; vorsichtiger *Stasson/Bradshaw*, Small Grp Res 1995, 296: "level of analysis was found to be an important consideration".

[496] *Michaelsen/Watson u.a.*, J Appl Psy 1992, 106, 107 berichten anekdotische Evidenz aus der Erinnerung der Kursleiter: "Students in these classes frequently expressed considerable pride in the fact that their group was able to answer a question correctly when none of the members had selected the preferred answer during the individual test."

[497] *Tindale/Larson*, J Appl Psy 1992, 102, 103.

[498] In ihrer Duplik halten *Michaelsen/Watson u.a.*, J Appl Psy 1992, 106 die Computersimulation für "suspect" und "very close to irrelevant" (106), und die Replikationsstudie wegen ihrer geringen Stichprobengröße für unbeachtlich ("they make so much of so little", 107); das geht an der Kritik freilich völlig vorbei.

[499] *Tindale/Larson*, J Appl Psy 1992, 102, 105 (Tabelle 1); *Larson*, Synergy 2010, 37.

[500] *Michaelsen/Watson u.a.*, J Appl Psy 1992, 106, 108: "most critical decisions in organizational settings […] involve the simultaneous consideration of a large number of factors […] Thus, using multifaceted problems, including total test scores […] seems to be a more valid way".

[501] So auch, nur anders akzentuiert, *Stasson/Bradshaw*, Small Grp Res 1995, 296, 298.

jede Form von Synergie. Eine solche – nämlich schwache Synergie – belegen sie auch ganz zweifelsfrei, allerdings ist das keineswegs überraschend und wird regelmäßig auch in Laborstudien beobachtet,[502] deren Übertragbarkeit die Autoren mit ihrer Studie ja anfangs in Frage gestellt hatten. Die Studie illustriert also auch die Gefahren des oftmals voreilig erhobenen Vorwurfs, die Ergebnisse von kontrollierten Experimenten seien Artefakte der Laborumgebung und sagten über „die Realität“ nichts aus. Deshalb fragte schon die erste Replik in ihrem Titel rhetorisch, ob die Autoren der Ursprungsstudie nicht einfach nur die „typische Leistung von Gruppen“ beobachtet hätten.[503] Diese Frage ist zu bejahen.

Abschließend ist darauf hinzuweisen, dass immerhin die zweite später durchgeführte Replikationsstudie Anhaltspunkte für einen Gruppierungsbonus erbringen konnte. Dort wählten zwar ebenfalls viele Gruppen (35 %) die falsche Lösung, obwohl ein Mitglied vorher schon die richtige gewusst hatte – aber fast genauso viele (28 %) wählten die richtige Lösung, obwohl *kein* Mitglied sie vorher kannte.[504] Dieser Unterschied zur ersten Replikationsstudie mag an deren geringer Stichprobengröße liegen, oder an der Aufgabenart der zweiten Replikationsstudie, die von ihren Versuchsteilnehmern allein mathematische Intelligenz erforderte.[505] Jedenfalls bieten sie den in der gesamten referierten Debatte zwischen 1989 und 1995 deutlichsten Beleg für echte Gruppierungsboni.

III. Entschleunigung der Informationssuche?

Die dritte Annahme zum gesellschaftsrechtlichen Kollegialprinzip lautet, dass es vorschnelle Entscheidungen durch sorgfältigere Informationssuche vermeidet (vgl. oben bei Fn. 295). Gruppenarbeit wird also als vorteilhaft gesehen, weil Gruppen mehr Zeit in die Informationssuche investieren. Andererseits gilt es gerade auch als *Nachteil* des Kollegialprinzips, dass Entscheidungen „namentlich in Pionierphasen und Krisenzeiten nicht so rasch getroffen werden [können] wie bei einer direktorialen Führungsstruktur.“[506]

---

[502] Vgl. *Stasson/Bradshaw*, Small Grp Res 1995, 296, 304: “data from an [sic] laboratory experiment in an artificial context showed nearly the identical effect size as a field study in which groups had significant rewards contingent on quality of performance.”

[503] *Tindale/Larson*, J Appl Psy 1992, 102: “Assembly Bonus Effect or Typical Group Performance?”

[504] *Stasson/Bradshaw*, Small Grp Res 1995, 296, 302 f.

[505] *Stasson/Bradshaw*, Small Grp Res 1995, 296, 301: “The data presented here were based on responses to a set of five items testing different mathematical concepts”.

[506] *Fleischer*, NZG 2003, 449, 458; *Roth Pellanda*, Organisation 2007, 209 (Rn. 426): „Entscheidfindungen benötigen erheblich mehr Zeit.“; *Wettich*, Vorstandsorganisation 2008, 301: „langwieriger Willensbildungsprozess“.

Das formulierte vor einigen Jahren der scheidende stellvertretende Vorsitzende des Vorstands der US-Notenbank so:

„Meine Tätigkeit im FOMC [Offenmarktausschuss der Notenbank] rief mir lebhaft einige Dinge ins Gedächtnis, die wohl jedermann über Gremien weiß: dass sie mühevoll individuelle Präferenzen aggregieren; dass sie der Führung bedürfen; dass sie in schwierigen Fragen Kompromisspositionen beziehen und – vielleicht als Konsequenz all dieser Punkte – dass sie träge sind.“[507]

Oberflächlich betrachtet scheint das Gesellschaftsrecht die Entscheidungsgeschwindigkeit von Kollegialorganen schlicht als Medaille mit zwei Seiten zu betrachten: „träge“ einerseits, aber zumindest nicht „vorschnell“ andererseits. Bei genauerem Hinsehen liegt der festgestellten „Trägheit“ allerdings ein anderer Bezugspunkt zugrunde als den unerwünschten „vorschnellen“ Entscheidungen:

Die Trägheit, die die beiden letzten Zitate unterstellen, bezieht sich auf den *Zeitaufwand* bei Kollegialentscheidungen. Dass Gruppen für Entscheidungen mehr Zeit durch Informationsaustausch und Diskussion aufwenden, ist eine verbreitete und fast nie bezweifelte Annahme.[508] Studien zum Problemlöseverhalten beobachten zwar bisweilen das Gegenteil – Gruppen fänden eine Lösung schneller und bräuchten nur nach Köpfen mehr Zeit als Individuen[509] – allerdings bedürften die Kommunikationsstrukturen dieser Studien und die Vergleichbarkeit ihrer Problemstellungen mit unternehmerischen Entscheidungen (vgl. bei Fn. 615) näherer Untersuchung. Ohne Weiteres ist die Vermutung, dass Kollegialorgane mehr Zeit aufwenden, nicht von der Hand zu weisen.

Allerdings kommt es für die vorliegende Frage gar nicht auf die Zeitdauer an, sondern auf das *Beharrungsvermögen*. Die gesetzgeberische Intention, „vorschnellen“ Entscheidungen vorzubeugen, beruht ersichtlich auf der Annahme, dass Kollegialorgane schwerer von einem bestimmten Standpunkt zu überzeugen seien und für die Entscheidungsfindung mehr Informationen berücksichtigen.[510] Wirtschaftliche Entscheidungen werden

---

[507] So *Alan Blinder* 1998, zit. in: *Blinder/Morgan*, J Mon Cred Bank 2005, 789, 790.

[508] So schon oben Fn. 231 und 506; vgl. aber auch *Grundei*, BCCG DiskP 2004, 10 f.: „im Grenzbereich geringer Entscheidungsqualitäten“ sei das Kollegium langsamer, ermögliche aber ansonsten eine „Verkürzung der zur Entscheidungsfindung erforderlichen Zeit.“; *Schmidt* in: Verwaltungsgerichtsordnung 1970, 85, 94: „was die kollegiale Beratung an Zeit verlieren lässt, bringt eben diese Beratung meist auch wieder ein.“

[509] *Baron/Kerr*, Group Process 2003, 38 („nach Köpfen“ meint, dass das Produkt aus Zeitverbrauch und Beteiligtenanzahl bei Gruppen zumeist größer ausfällt als bei Individuen); *Hill*, Psy Bull 1982, 517, 524.

[510] Zur Verknüpfung von Zeitbedarf und Fundierungsintensität *Grundei*, BCCG DiskP 2004, 1/10 f.; für die Verwaltungsorganisation *Dagtoglou*, Kollegialorgane 1960: „Dieser langsamere Arbeitsgang ist nicht nur ein rein zeitlicher. Ein Kollegialorgan ist nicht nur weniger prompt und wendig […] Vielmehr besitzt es eine naturgemäße *konservative* Neigung“.

– mit Ausnahme des Echtzeithandels – oft weniger nach der insgesamt benötigten Zeitdauer beurteilt als nach der Verzögerung gegenüber neuen Entwicklungen oder Informationen.[511] Auch wochenlang geplante, aber schlecht informierte Entscheidungen sind also „vorschnell" – ebenso wie wohlreflektierte Entscheidungen auch dann nicht „vorschnell" sind, wenn sie binnen einer halben Stunde gefasst werden.

Dass Kollegialorgane in diesem letzten Sinn weniger vorschnell handeln als Einzelentscheider, ist nun gar nicht so selbstverständlich, wie es zunächst scheint. So findet sich im älteren gesellschaftsrechtlichen Schrifttum sogar die Annahme, dass ergebnisorientierte Gruppen *zu schnell* entscheiden, weil sie „eine einstimmige, aber falsche, Entscheidung schon sehr früh im Entscheidungsprozess treffen."[512] Dies beruhe darauf, dass Gruppen ihre Informationssuche auf das Nötigste beschränkten und nur so lange diskutierten, bis die erste konsensfähige Lösung erreicht sei.[513]

In der empirischen Literatur finden sich vor allem zwei einschlägige Experimente, die der eingangs zitierte Notenbankvertreter nach seiner Rückkehr in die Wissenschaft durchführte.[514] Eines der beiden Experimente war sehr abstrakt, das andere unmittelbar auf Geldmarktpolitik zugeschnitten und entsprechend kompliziert im Aufbau. Da beide fast identische Ergebnisse erbrachten, sei nur der Aufbau des einfacheren Versuchs geschildert:[515] Teilnehmern wurde eine virtuelle Urne vorgesetzt, die gleich viele blaue und rote Bälle enthielt. Zudem wurde ihnen mitgeteilt, dass sich diese Zusammensetzung im Laufe des Experiments genau einmal ändern würde. Ihre Aufgabe bestand darin, durch Zufallsziehungen mit Zurücklegen herauszufinden, ob die Urne am Ende des Experiments mehr rote oder mehr blaue Bälle enthielt. Bekannt war lediglich, dass die Zusammensetzung der Urne nach einer der ersten zehn Ziehungen (mit jeweils gleicher Wahrscheinlichkeit) geändert werden würde und dass danach entweder 70 % blaue und 30 % rote Bälle enthalten sein würden, oder 70 % rote und 30 % blaue Bälle.[516] Die Teilnehmer konnten bis zu vierzig Mal einen Ball ziehen und mussten nach der Änderung der Urnenzusammensetzung für je-

---

[511] So *Blinder/Morgan*, J Mon Cred Bank 2005, 789, 794 f.; a.A. *Baron/Kerr*, Group Process 2003, 38: "an employer's payroll ultimately depends on the number of person-minutes required to complete some task".

[512] *Haft*, Mich L Rev 1981, 1, 37 ff., und weiter: "Contrary to folklore, […] the real difficulty is trying to slow groups down rather than trying to speed them up", m.Verw. auf *Barnlund*, J Abnorm Soc Psy 1959, 55, 59.

[513] *Haft*, Mich L Rev 1981, 1, 39: "Groups adopt only as many criteria as are needed to solve the problem, and the group's discussion centers around the first apparently reasonable solution. If enough cues reveal that a solution is acceptable, it becomes the group decision."

[514] *Blinder/Morgan*, J Mon Cred Bank 2005, 789; *Blinder/Morgan*, Int J Central Bank 2008, 117.

[515] Nach *Blinder/Morgan*, J Mon Cred Bank 2005, 789, 803 ff.

den gezogenen Ball ein Vierzigstel ihres Anfangsguthabens bezahlen. Sie konnten zu einem beliebigen Zeitpunkt einmal raten, ob die Urne mehr blaue oder mehr rote Bälle enthielt. Hatten sie die neue Zusammensetzung richtig erraten, gewannen die Teilnehmer das anderthalbfache ihres Anfangsguthabens hinzu. Rieten sie falsch oder zu früh (also schon vor der Änderung), erhielten sie nichts – und mussten, wenn sie zu früh geraten hatten, außerdem ein Vierzigstel ihres Anfangsguthabens für jede Runde bezahlen, die sie zu früh geraten hatten. Dieser Aufbau sollte Situationen nachbilden, die eine Abwägung erfordern zwischen einer fehleranfälligen Entscheidung unter Unsicherheit und kostspieligem Abwarten zur weiteren Informationssuche.

Die Teilnehmer des Experiments entschieden zunächst einige Male allein, anschließend in einer Fünfergruppe, in der sie frei kommunizieren konnten, und zuletzt noch einmal allein. Überraschenderweise ergab das Experiment keinen Anhaltspunkt dafür, dass die Gruppe langsamer[517] entschied als ihre einzelnen Mitglieder. Zwar benötigten Gruppen mit durchschnittlich 6,6 nach dem Farbwechsel gezogenen Bällen etwas länger als Individuen mit durchschnittlich 6,4 nach dem Farbwechsel gezogenen Bällen – allerdings war dieser Unterschied trotz der hohen Stichprobengröße nicht signifikant, kann also nicht auf eine systematische Ursache zurückgeführt werden. Dieses Ergebnis lässt daran zweifeln, dass Gruppen tatsächlich länger zögern und mehr Informationen sammeln, bevor sie eine Entscheidung fällen. Zumindest wenn Abwarten finanzielle Kosten mit sich bringt, scheinen Gruppen ebenso schnell zu entscheiden wie Individuen.

Mehrere Argumente sprechen dafür, dass dieser Befund auf gesellschaftsrechtliche Zusammenhänge übertragbar sein könnte: Erstens erbrachte das andere (komplexere und noch deutlicher wirtschaftsbezogene) Experiment derselben Autoren das gleiche Ergebnis[518] und konnte mit anderen Teilnehmern repliziert werden.[519] Zweitens hatten die Autoren den Teilnehmerkreis ihrer Versuche von vornherein auf makroökonomisch vorgebildete Studenten beschränkt,[520] denen die in den Versuchen geschilderten Szenarien durchaus vertraut gewesen sein dürften.

---

[516] Nicht mitgeteilt wurde, wie diese Änderung stattfinden würde. Statistisch sind die denkbaren Varianten – Entnahme, Umfärbung oder Hinzugabe von Bällen – gleichwertig, allerdings ist wohl davon auszugehen, dass die Teilnehmer ganz verschiedene mentale Repräsentation des Problems gebildet haben.

[517] *Blinder/Morgan*, J Mon Cred Bank 2005, 789, 795: “when we say that one type of decisionmaking process ‘takes longer’ than another, we mean that more data (not more minutes) are required before the decision is made.”

[518] *Blinder/Morgan*, J Mon Cred Bank 2005, 789, 798.

[519] *Blinder/Morgan*, Int J Central Bank 2008, 117.

[520] *Blinder/Morgan*, J Mon Cred Bank 2005, 789, 792; *Blinder/Morgan*, Int J Central Bank 2008, 117, 121.

IV. Gegenseitige Überwachung und Mäßigung?

Die vierte Annahme zum gesellschaftsrechtlichen Kollegialprinzip lautet, dass es die Überwachungseffizienz erhöht und durch interne Kompromissbildung zu ausgewogeneren Entscheidungen führt (vgl. oben bei Fn. 296). Für diese Annahme sind vor allem drei Phänomene aus der Gruppenforschung von Relevanz.

1. Gruppendenken (groupthink)

Zweifel an der Verbesserung der Überwachungseffizienz werden schon im gesellschaftsrechtlichen Schrifttum selbst laut. Als einer der wichtigsten Effekte von Gruppenentscheidungen allgemein[521] und im Kontext der Unternehmensführung insbesondere[522] gilt nämlich „ein Phänomen, das unter der Bezeichnung ‚Groupthink' gerade im Kontext hochrangiger (politischer und wirtschaftlicher) Entscheidungen bekannt geworden ist" und Zweifel an der Überwachungseffizienz von Kollegialorganen begründe.[523] Solches „Gruppendenken" sei in der Theorie

> „dadurch gekennzeichnet, dass für die Mitglieder einer hochkohäsiven Gruppe das Streben nach Einmütigkeit eine höhere Bedeutung gewinnt als die realistische Einschätzung von Entscheidungsalternativen. Groupthink äußert sich konkret beispielsweise in einer Illusion der Unverwundbarkeit, die übertriebenen Optimismus und einen Hang zum Risiko nach sich zieht, in der Geringschätzung von Warnungen bzw. unpassenden Informationen, in der Ausübung von Uniformitätsdruck auf (potentiell) deviante Gruppenmitglieder sowie in einer Illusion einmütiger Meinungen. Unter diesen Bedingungen ist eine mangelhafte Entscheidungsqualität zu erwarten, die vornehmlich darin zum Ausdruck kommt, dass insgesamt eine unzureichende, verzerrte Informationssuche und -verarbeitung stattfindet, berücksichtigte Alternativen nicht noch einmal überdacht und keine Alternativpläne erstellt werden und die Risiken der gewählten Alternative nicht untersucht werden. Zeitdruck zählt dabei zu den Situationsbedingungen, die eine unzureichende Entscheidungsfundierung begünstigen."[524]

---

[521] *Baron*, Adv Exp Soc Psy 2005, 219: "arguably the most widely publicized application of psychological principles to high-level […] group decision making in the history of experimental psychology".

[522] *Barnard*, N Car L Rev 1991, 1135, 1171 ff.; *Bainbridge*, Vand L Rev 2002, 1, 32: "most significant group bias" im Kontext von Vorstandsentscheidungen; *Dorff*, Cardozo L Rev 2007, 2025, 2035 ff.

[523] *Grundei/von Werder*, AG 2005, 825, 829; *Langevoort*, U Penn L Rev 1997, 101, 138; *Fanto*, Or L Rev 2004, 435, 462, je m.Verw. auf *Janis*, Groupthink 1972; pointiert *Haft*, Mich L Rev 1981, 1, 29: es bedürfe einer "fine balance between incest and indifference among group members."

[524] *Grundei/von Werder*, AG 2005, 825, 829 in enger Anlehnung an die ursprüngliche Formulierung des Konzepts durch *Janis*, Groupthink 1972, 8 f.; vgl. auch *Baron*, Adv Exp Soc Psy 2005, 219, 220 f.; speziell zur Unternehmensführung *Bainbridge*, Vand L Rev 2002, 1, 32; *Grundei*, BCCG DiskP 2004, 1/13. Die fast gegenteilige Verwendung des Begriffs in *Europäische Kommission*, Grünbuch 2011, 6 beruht auf einem Übersetzungsfehler: Wo es in der

Dementsprechend gilt Gruppendenken oft als Grund für unternehmerische Fehlentscheidungen bis hin zum systematischen Betrug,[525] bisweilen sogar im weltwirtschaftlichen Maßstab wie im Vorfeld der Enron- und WorldCom-Insolvenzen.[526] Gruppendenken könnte also die Überwachungseffizienz des Kollegialorgans empfindlich vermindern.[527] Andererseits könnte Gruppendenken seinerseits eine Folge mangelnder Kollegialität sein,[528] gilt es doch im Vorstand als umso wahrscheinlicher, je mächtiger der Vorstandsvorsitzende ist, weil dieser versuche, seine Entscheidungsmacht zu behaupten, Gefolgschaft zu belohnen und Abweichler auszubooten.[529] Womöglich führt das Kollegialprinzip also gar nicht zum Gruppendenken – sondern hegt es im Gegenteil geradezu ein – wenn es die Gleichberechtigung der Organmitglieder konsequent durchführt. Jedenfalls fragt sich, welche empirischen Belege die Theorie des Gruppendenkens stützen.

Angesichts seiner Verbreitung als Argumentationstopos, seiner feinen theoretischen Ausarbeitung und der ihm beigemessenen praktischen Bedeutung existiert bislang überraschend wenig empirische Substanz, die die Hypothesen über das Gruppendenken stützt – zu diesem Ergebnis kam 2005 eine umfassende Sichtung empirischer Befunde und früherer Literaturübersichten.[530] Obwohl schon das ursprünglich von Irving Janis formulierte Modell spezifische Beziehungen zwischen genau benannten Einflussfaktoren postulierte und damit für die empirische Überprüfung bestens geeignet gewesen sei, fänden sich zu den meisten Einflussfaktoren tatsächlich nur wenige Studien, die überdies noch widersprüchliche Ergebnisse aufwiesen.[531] Die Gesamtschau führt zu einem ernüchternden Fazit:

---

deutschen Fassung „zum ‚Gruppendenken' beitragen" heißt, steht in der englischen Fassung "to tackle 'group-think'" (engl. *tackle:* bewältigen, in Angriff nehmen).

[525] *Neck*, J Manage Psy 1996, 3 zu Fehlentscheidungen bei NBC; *Sims*, J Bus Ethics 1992, 651 zu Betrugsfällen beim Babynahrungsproduzenten Beech-Nut und den Finanzdienstleistern E. F. Hutton und Salomon Brothers.

[526] *O'Connor*, U Cin L Rev 2003, 1233 zu Enron; *Scharff*, J Lead Org Stud 2005, 109 zu WorldCom; vgl. auch *Bénabou*, NBER (14764) 2009, 3, 28 (Appendix A) mit Belegen "from NASA to the FED, SEC and Fannie Mae, from Enron to major investment banks, A.I.G and individual investors."

[527] Anders offenbar *Bainbridge*, Vand L Rev 2002, 1, 32: "To the extent groupthink promotes the development of social norms, it facilitates the board's monitoring function."

[528] So *Seidenfeld*, Cornell L Rev 2002, 486, 543: "I suspect that groupthink is more of a problem when the work group's interactions become hierarchical."

[529] *Bainbridge*, Vand L Rev 2002, 1, 32: "Boardroom culture encourages groupthink. […] CEOs foster and channel groupthink through the exercise of their powers to control information flows, reward consensus, and discourage reelection of troublemakers."; auch *Baron/Kerr*, Group Process 2003, 98 sehen Gruppendenken oft durch den Gruppenleiter verursacht.

[530] *Baron*, Adv Exp Soc Psy 2005, 219; ebenso *Kugler/Kausel/Kocher*, WIREs Cog Sci 2012, 471, 473: "findings supporting groupthink are sparse."

„Auf der Grundlage dieser enttäuschend mangelhaften Untermauerung seiner Hauptvorhersagen empfiehlt die große Mehrheit der Literaturübersichten eine Überarbeitung, Ersetzung oder sogar die glatte Aufgabe dieses Modells.“[532]

Mit der Schlussfolgerung, das Modell des Gruppendenkens sei empirisch schwach untermauert, ist für die vorliegende Fragestellung allerdings noch nicht viel gewonnen. Die Unzulänglichkeiten realer Entscheidungsprozesse, die das Modell erklären sollte, bestehen ja auch dann, wenn das Modell sie nicht hinreichend erklären kann. Schon die enorme Verbreitung des Topos' „Gruppendenken“ spricht ja dafür, dass er in den Lebenserfahrungen der meisten Menschen eine hohe Plausibilität beanspruchen kann.[533] Dementsprechend berufen sich alternative Erklärungsversuche weniger auf die von der Theorie des Gruppendenkens postulierten Einflussfaktoren (Gruppenkohäsion, Identitätsbedrohung, usw.) als vielmehr auf universelle psychologische Gruppendynamiken.[534] Die bisherige Forschung über das Gruppendenken mündete also langfristig weniger in die Untermauerung dieser Theorie als in kontroverse Arbeiten über die Kehrseite von Gruppenentscheidungen.[535] Eine solche Kehrseite, die handfeste Zweifel am Beitrag des Kollegialprinzips zur Überwachungseffizienz aufwerfen und gleichzeitig dazu beitragen kann, unternehmerische Fehlentscheidungen zu verstehen, ist die sog. Gruppenpolarisierung, die der Idee vom Gruppendenken ähnelt, aber keine besonders kohäsive oder sozial isolierte Gruppe voraussetzt.[536]

2. Gruppenpolarisierung (group polarization)

Die vermeintliche Ausgewogenheit kollegialer Entscheidungen wurde oben maßgeblich auf die „Tendenz der Gruppe“ zurückgeführt, „extreme Urteile zu vermeiden und sich auf mittlere Linien zu beschränken.“[537] Während die

---

[531] *Baron*, Adv Exp Soc Psy 2005, 219, 222 f. m.w.N.; vgl. schon *Levine/Moreland*, Ann Rev Psy 1990, 585, 604, 619: “the theory has certain conceptual limitations and […] weak research support”.

[532] *Baron*, Adv Exp Soc Psy 2005, 219, 224 m.w.N.; krit. schon *Seidenfeld*, Cornell L Rev 2002, 486, 541 f.

[533] *Baron*, Adv Exp Soc Psy 2005, 219, 227: “this model has a certain ‘ring of truth’ that resonates with readers”.

[534] *Baron/Kerr*, Group Process 2003, 107 f.; *Baron*, Adv Exp Soc Psy 2005, 219, 228 ff.: “the ubiquity of groupthink” erkläre sich aus “the general prevalence of consensus seeking, group polarization, outgroup stereotyping, and the suppression of dissent”.

[535] *Kerr/Tindale*, Ann Rev Psy 2004, 623, 640.

[536] Genau umgekehrt argumentiert *Sibert*, Int Fin 2006, 145, 162, die zunächst die Gruppenpolarisierung einführt und dann das Gruppendenken als “a particular type of group polarization” – allerdings ohne zu reflektieren, ob „Gruppendenken“ einen über die Gruppenpolarisierung hinausreichenden Erklärungswert hat.

Sozialpsychologie dieses Verständnis ursprünglich teilte[538] und die Ökonomik ihm noch heute zuneigt,[539] hat die empirische Forschung daran erhebliche Zweifel aufkommen lassen, die vereinzelt auch im gesellschaftsrechtlichen Schrifttum aufgegriffen wurden.[540]

In seiner unveröffentlichten Masterarbeit beobachtete James Stoner 1961 etwas, das zunächst als Risikoverschiebung (*risky shift*) in die Verhaltenswissenschaft einging: Dieselben Individuen verhielten sich deutlich risikofreudiger, wenn sie Entscheidungen nach einer Diskussion als Gruppe trafen als wenn sie die gleichen Entscheidungen allein und unabhängig voneinander zu treffen hatten.[541] Seither haben hunderte Studien gezeigt,[542] dass Gruppen ganz allgemein die moderaten Einstellungen ihrer Mitglieder in Richtung des Extrems polarisieren (*group-induced attitude polarization*): risikofreudige Individuen werden in der Gruppe risikofreudiger, risikoscheue Individuen bilden noch risikoscheuere Gruppen, und feministische Frauen werden nach einer Gruppendiskussion feministischer.[543] Diese Art der Polarisierung meint also nicht – wie der Begriff auch verstanden werden könnte[544] – ein Auseinanderdriften der Einstellungen innerhalb der Gruppe, sondern ein gemeinsames Driften der Gruppe in Richtung ihres Extrems. Wesentliche Voraussetzung für diese polarisierende Wirkung der Gruppenberatung ist die vorherige Existenz eines grundsätzlichen Gruppenkonsenses; zur Risikoverschiebung etwa kommt es nur, wenn schon vor

---

537 *Berkemann*, JZ 1971, 537, 539, zit. schon oben bei und in Fn. 258; dagegen verweist *Berkemann*, KritV 1988, 29 ausdrücklich auf den Polarisierungseffekt in Kleingruppen (dazu gleich im Haupttext).

538 *Baron/Kerr*, Group Process 2003, 98: “Initially, researchers assumed that group discussion would have a mellowing influence on hotheads and extremists within the group.”

539 *Sibert*, Int Fin 2006, 145, 147: Moderation durch Gruppen “would appear to be the view from the economics literature.”

540 *Roth Pellanda*, Organisation 2007, 209 (Rn. 425): „Nach empirischen Studien sind Gruppenentscheidungen [.] oftmals auch risikofreudiger als die durch eine Person getroffene Entscheidung.“; *Fanto*, Or L Rev 2004, 435, 465; *Sharfman/Toll*, Nw U L Rev Coll 2009, 380, 388 f.

541 *Isenberg*, J Pers Soc Psy 1986, 1141 m.w.N.

542 *Baron/Kerr*, Group Process 2003, 103: “over 500 studies”; *Sunstein*, NYU L Rev 2005, 962, 1004: “hundreds of studies involving over a dozen countries”; vorsichtiger *Sibert*, Int Fin 2006, 145, 161: “more than 300 studies”; schon *Isenberg*, J Pers Soc Psy 1986, 1141 erwähnte “several hundred studies”.

543 Bsp. nach *Isenberg*, J Pers Soc Psy 1986, 1141 m.w.N.; weitere Bsp. bei *Baron/Kerr*, Group Process 2003, 98; krit. *Petersen*, Staat 2010, 435, 448 (ohne Bezug auf Forschungssynthesen oder neuere Literatur): „dass in bestimmten Experimenten herausgefunden wurde, dass Deliberation unter bestimmten Umständen zur Polarisierung der Diskussionsteilnehmer führt, [bedeutet] nicht, dass Deliberation *per se* zu einer Polarisierung führt.“

544 So etwa bei *Schulz Von Thun*, Miteinander reden 2006, 240.

der Diskussion die Mehrheit der Gruppenmitglieder zur Eingehung von Risiken in gewissem Umfang bereit ist.[545]

Hat die Gruppe anfänglich bereits einen gewissen, aber unausgesprochenen Grad an Übereinstimmung, sorgen vor allem zwei Prozesse für eine Polarisierung dieser Übereinstimmung hin zum Extrem[546] – der Wettbewerb mit anderen Gruppenmitgliedern (*social comparison*) und der Austausch überzeugender Argumente (*persuasive argumentation*):[547] Zum einen wetteifern Gruppenmitglieder unbewusst um soziale Zugehörigkeit und versuchen deshalb bei einer als Konsens empfundenen Position, leicht überdurchschnittliche Gruppenloyalität an den Tag zu legen, ohne dass es einer Gruppendiskussion überhaupt bedürfte; zum anderen werden geteilte Meinungen in der Gruppendiskussion häufiger ausgesprochen und mit höherer Wahrscheinlichkeit wiederholt, was jeweils das Vertrauen in ihre Richtigkeit erhöht und Gruppenmitglieder entweder davon überzeugen oder darin bestärken kann, dass der anfänglich moderate Gruppenkonsens richtig ist.[548] Dass keiner der beiden Prozesse den Polarisierungseffekt allein erklären kann, sondern beide in der Praxis zusammenwirken, haben unabhängig voneinander sowohl eine Metastudie über 21 Primärstudien (1974–1982) als auch ein aktuellerer Forschungsbericht von 2003 ergeben.[549]

3. Einsatzeskalation (escalation of commitment)

Ein im Zusammenspiel mit der Gruppenpolarisierung für das Gesellschaftsrecht besonders relevantes Verhaltensphänomen ist die sog. Einsatzeskalation (*escalation of commitment*).[550] Damit wird die Tendenz von Entscheidungsträgern bezeichnet, verlustträchtige Entscheidungen dadurch psychologisch zu rechtfertigen, dass sie ihre Investition noch erhöhen – also „dem schlechten Geld gutes hinterherwerfen“ in der Hoffnung, ihre frühere Entscheidung erweise sich doch noch als gut und profitabel. Diese

---

[545] *Baron/Kerr*, Group Process 2003, 98 f.

[546] So die seit Ende der 1970er Jahre vorherrschende Dichotomie, vgl. *Isenberg*, J Pers Soc Psy 1986, 1141 m.w.N. zur früheren Forschung; *Baron/Kerr*, Group Process 2003, 99 f., auf 100 ff. mit einem dritten Erklärungsansatz (*self-categorization*), der noch nicht hinreichend etabliert ist, um ihn hier gewinnbringend zu besprechen.

[547] Die nachfolgenden Beschreibungen der beiden Prozesse folgen *Baron/Kerr*, Group Process 2003, 99 f.; ähnl. auch *Sunstein*, Yale L J 2000, 71, 88–90; *Sunstein*, J Polit Phil 2002, 175, 176–180; *Seidenfeld*, Cornell L Rev 2002, 486, 535–537.

[548] Hier besteht eine Wechselbeziehung zum Problem verborgener Informationsprofile, vgl. oben bei Fn. 448.

[549] *Isenberg*, J Pers Soc Psy 1986, 1141, 1149: “outside of the laboratory they almost always co-occur”; *Baron/Kerr*, Group Process 2003, 99 f.

[550] *Langevoort*, U Penn L Rev 1997, 101, 142 f.; *Langevoort*, Geo L J 2001, 797, 811; *Paredes*, Fla St U L Rev 2005, 673, 691.

von Barry Staw erstmals aufgezeigte Urteilsverzerrung hängt maßgeblich davon ab, ob der Entscheidungsträger, der die zweite Investition beurteilt, auch für die erste Investition verantwortlich war – oder sich dafür verantwortlich fühlt.[551] Hier nun kommt die Gruppenpolarisierung zum Tragen, die das Vertrauen aller Gruppenmitglieder bestärkt, gemeinsam die richtige Entscheidung getroffen zu haben. Die möglichen Vorteile „einer besseren Identifikation sämtlicher Vorstandsmitglieder mit der Entscheidung“[552] bergen also auf ihrer Kehrseite die Gefahr, dass sämtliche Vorstandsmitglieder unvernünftig lange an verlustträchtigen Entscheidungen festhalten.[553]

In der Praxis wurde zunächst die industrielle Überproduktion der 1980er Jahre auf die Einsatzeskalation zurückgeführt.[554] Zudem gilt Einsatzeskalation als mittelbare Ursache für die geringe Profitabilität US-amerikanischer Unternehmenskäufe in den 1980er Jahren,[555] denn sie befördere aggressive Bieterwettbewerbe (*bidding wars*) wie 1987/88 um die Kaufhauskette Bloomingdale's, deren feindliche Übernahme gegen den Widerstand eines Konkurrenten den Erwerber geradewegs in die Insolvenz führte.[556] Als jüngeres Beispiel gilt der Medizinproduktehersteller Guidant, dessen Übernahme 2005/06 den „siegreichen“ der beiden konkurrierenden Bieter mehr als die Hälfte seiner Marktkapitalisierung gekostet hatte.[557] Auch das Verhalten von treulosen Wertpapierhändlern (*rogue traders*) wie Nick Leeson, der 1995 den Zusammenbruch der ältesten Geschäftsbank Londons verursachte, „demonstriert die dramatischen Konsequenzen, die die Eskalation von Einsätzen zur Vermeidung von Verlusten haben kann“.[558] Angesichts dieser praktischen Relevanz stellt sich die Frage, wie kollegiale Gruppenarbeit die Neigung zur Einsatzeskalation beeinflusst.

Zwei Vignettenstudien (dazu oben § 3 B.IV.2.c) von 1984 bzw. 1993 untersuchten den Zusammenhang von Einsatzeskalation und Gruppenarbeit.[559] Die erste konnte belegen, dass Gruppen von der Einsatzeskalation ebenso betroffen sind wie Individuen.[560] Da sie aber keinen direkten Ver-

---

[551] *Bazerman/Giuliano/Appelman*, Org Behav Hum Perf 1984, 141, 142 f. m.w.N.

[552] *Wettich*, Vorstandsorganisation 2008, 300.

[553] *Bazerman/Giuliano/Appelman*, Org Behav Hum Perf 1984, 141, 143 f., aber unter Rückgriff auf die inzwischen überholten Konzepte Gruppendenken und Risikoverschiebung.

[554] *Langevoort*, U Penn L Rev 1997, 101, 143 mit zwei Praxisbeispielen in Fn. 147.

[555] *Bazerman/Moore*, Judgment 2009, 105: “in the aggregate, the synergy that was obtained in acquisitions went to the sellers”.

[556] *Bazerman/Neale*, Eur Mgmt J 1992, 163, 163 f.

[557] *Bazerman/Moore*, Judgment 2009, 107 f. mit dem Gegenbsp. von American Airlines, die 1995 einen Bieterwettbewerb mit United Airlines um den Erwerb der USAir verhinderten.

[558] *Bazerman/Moore*, Judgment 2009, 109 f.

[559] Zu älteren Studien ausf. *Brockner/Rubin*, Escalating Conflicts 1985, 57 ff. (Kap. 5).

[560] *Bazerman/Giuliano/Appelman*, Org Behav Hum Perf 1984, 141, 150; *Bazerman/Moore*, Judgment 2009, 104 entnehmen dieser Studie “that groups are less likely than individuals

gleich zwischen Gruppen und Individuen berichtete, lässt sich dieser Studie nicht entnehmen, ob die Einsatzeskalation von Gruppen stärker oder weniger stark ausfällt als jene von Individuen.[561] Die zweite Vignettenstudie ist insoweit aufschlussreicher. Ihre Teilnehmer mussten sich zuerst allein, danach in Fünfergruppen[562] je dreimal in die Rolle von Entscheidungsträgern versetzen, die verschiedene Investitionsentscheidungen zu treffen haben.[563] In einer der drei Vignetten war über die Erstinvestition zu entscheiden, in den beiden anderen über Folgeinvestitionen, nachdem die Erstinvestition – die der Entscheidungsträger in einer Vignette auch noch selbst zu verantworten hatte – erfolglos aufgebraucht worden war.[564] In allen Fällen erhielten die Teilnehmer finanzielle Prognosen, aus denen sich die wirtschaftliche Neutralität der zu beschließenden Investition ergab (Erwartungswert null). Dadurch konnten die Antworten auf Vignetten mit und ohne persönlicher Verantwortlichkeit verglichen werden, um das Ausmaß der Einsatzeskalation zu bestimmen.

Im Ergebnis befürworteten 29 % der Individuen und 26 % der Gruppen die Erstinvestition und 66 % der Individuen sowie 77 % der Gruppen die Folgeinvestition; bei persönlicher Verantwortlichkeit für die Erstinvestition stieg der Zuspruch für die Folgeinvestition auf 72 % für Individuen und 94 % für Gruppen. Damit investierten also Gruppen ebenso wie Individuen öfter zum wiederholten Mal, wenn sie bereits für einen ersten Fehlschlag persönlich verantwortlich waren (Einsatzeskalation), doch fiel der Zuwachs bei den Gruppen noch deutlich stärker aus,[565] was auf ihre polarisierende Wirkung hinweisen könnte. Die Studie bietet also einen Anhaltspunkt dafür, dass Gruppen noch anfälliger für eine Einsatzeskalation sind als einzelne Entscheidungsträger.[566]

---

to escalate commitment; however, groups that escalate tend to do so to a greater degree than individuals."; das folgt aus der Originalveröffentlichung allerdings nicht.

[561] Tabelle 1 bei *Bazerman/Giuliano/Appelman*, Org Behav Hum Perf 1984, 141, 148 spricht stark für eine stärkere Einsatzeskalation: Folgeinvestitionen fallen bei Individuen 24,6 %, bei Gruppen dagegen 68,9 % höher aus, wenn sie für die Erstinvestition verantwortlich waren; das verkennt *Whyte*, Org Behav Hum Dec Proc 1993, 430, 432.

[562] Gruppen durften frei kommunizieren und entscheiden, *Whyte*, Org Behav Hum Dec Proc 1993, 430, 440.

[563] Im Einzelnen *Whyte*, Org Behav Hum Dec Proc 1993, 430, 439, ab 450 (Appendix 1) mit dem Volltext des Szenarios „Finanzvorstand eines großen Technologie-Unternehmens".

[564] *Whyte*, Org Behav Hum Dec Proc 1993, 430, 437 ff.

[565] *Whyte*, Org Behav Hum Dec Proc 1993, 430, 442: "significant interaction between decision frame and performing unit indicates that decision frame had a more pronounced effect on the frequency of escalation in group decision making than in individual decision making."

[566] Optimistischer *Sniezek*, Org Behav Hum Dec Proc 1992, 124, 147, die konkrete Interventionen bespricht und daraus schließt: "the group decision-making unit has a unique *potential* to 'debias' nonrational escalation of commitment" (Hervorhebung nur hier).

Freilich ist es auch hier wiederum angebracht, die Übertragbarkeit der Studien zu hinterfragen. Vignettenstudien sind ihrer Natur nach hypothetisch und konfrontieren Teilnehmer bisweilen sogar mit Szenarien, mit denen sie überhaupt nicht vertraut sind. Für die hier vorgestellten Studien spricht allerdings, dass ihre Entscheidungsszenarien gezielt auf wesentliche Unternehmensentscheidungen zugeschnitten waren und dass der Teilnehmerkreis der zweiten Studie demselben Bildungshintergrund entstammte wie Unternehmensleiter in der Realität.[567]

V. Verringerung der Selbstüberschätzung?

Die fünfte und letzte Annahme zum Kollegialprinzip lautet, dass es den Einfluss der urteilsverzerrenden Selbstüberschätzung reduziert (vgl. oben bei Fn. 297). Bisweilen wird diese Annahme mit anekdotischer Evidenz unterfüttert,[568] aber eine systematische Sichtung der empirischen Literatur steht auch hier noch aus. Bei der Untersuchung, welche Erkenntnisse die empirische Forschung über die genannte Annahme beisteuern kann, werde ich erneut in einem Zweischritt vorzugehen: Zunächst werde ich hinterfragen, ob Unternehmensleiter überhaupt der Selbstüberschätzung unterliegen und worin sich das äußert. Danach werde ich untersuchen, ob diese Selbstüberschätzung durch kollegiale Tätigkeit eingedämmt werden kann.

1. Selbstüberschätzung (overconfidence) bei Unternehmensleitern

Als Selbstüberschätzung wird eine kognitive Urteilsverzerrung bezeichnet, die ihren Ausdruck in einer ganzen Reihe verschieden akzentuierter Effekte findet – etwa in der als Überoptimismus (*overoptimism* bzw. *optimism bias*) bezeichneten Tendenz, mehr Erfolge als Misserfolge zu erwarten; in der selbstwertdienlichen Verzerrung (*self-serving bias*), Erfolge der eigenen Person, Misserfolge dagegen der Umwelt zuzuschreiben;[569] und in der als übersteigertes Selbstvertrauen (*overconfidence bias*) bezeichneten generellen Überschätzung des eigenen Könnens.[570] Die genaue Abgrenzung dieser Effekte ist nicht immer ganz eindeutig und kann hier dahinstehen.[571]

---

[567] *Whyte*, Org Behav Hum Dec Proc 1993, 430, 437: 200 Absolventen und 125 fortgeschrittene Studenten der Betriebswirtschaft (*business administration*).

[568] *Bainbridge*, Vand L Rev 2002, 1, 30 nennt “RJR Nabisco’s efforts to develop a smokeless cigarette” und ein für die gegenständliche Frage indirekt relevantes Experiment.

[569] Kulturübergreifende Metastudie über 266 Primärstudien in *Mezulis/Abramson u.a.*, Psy Bull 2004, 711; ältere Metastudie über 22 Feldstudien bei Sportereignissen bei *Mullen/Riordan*, J Appl Soc Psy 1988, 3.

[570] Dieselben oder ähnliche Konzepte kursieren auch unter den Bezeichnungen *above average effect*, *Lake Wobegon effect*, *illusory superiority*, *superiority bias* und *leniency error.*

[571] Ebenso *Langevoort* in: Gigerenzer/Engel, Heuristics and Law 2006, 87, 92.

Ähnlich wie große Teile der Allgemeinheit unterstellt das gesellschaftsrechtliche Schrifttum gern, dass Unternehmensleiter sich „bekanntermaßen“ systematisch selbst überschätzen.[572] Dagegen finden sich durchaus auch kritische Stimmen, die darauf hinweisen, dass Unternehmensleiter regelmäßig die „Überlebenden eines Wettbewerbs [um die Führungsposition] sind, der unterlegene kognitive Charakterzüge ausmerzen sollte“.[573] Da die ersten Studien zur Selbstüberschätzung aus Laborversuchen mit Studenten stammen, die einem solchen Wettbewerbsdruck nicht unterliegen, ist die Frage also nicht unberechtigt, ob Selbstüberschätzung auch „in realen Situationen existiert, oder ob sie nur ein Laborartefakt darstellt“.[574]

Die empirische Literatur ist in einer guten Position zur Beantwortung dieser Frage, denn Selbstüberschätzung ist die wohl am eingehendsten studierte Urteilsverzerrung unter Entscheidungsträgern im Unternehmen.[575] Sie wurde zwar auch für selbstständige Einzelunternehmer belegt,[576] aber der Großteil der Forschung bezieht sich auf *chief executive officer (CEO)* US-amerikanischen Zuschnitts. Da deren Stellung am ehesten der eines starken Vorstandsvorsitzenden oder gar Alleinvorstands entspricht, sollten die meisten Erkenntnisse ein unverfälschtes Bild der Selbstüberschätzung von Unternehmensleitern wiedergeben, wie sie sich ohne moderierende Gruppenprozesse darstellt. Der Stand der umfangreichen experimentellen Forschung lässt sich wie folgt zusammenfassen:

> „Viele Manager halten ihr Unternehmen, und sich selbst, für besser als den Durchschnitt. Ein Drittel von ihnen hält seine Misserfolgswahrscheinlichkeit glatt für null; über 80 % schätzen diese Wahrscheinlichkeit zwischen 0 und 30 %. Manager unterschätzen die Stärke ihrer Wettbewerber systematisch. […] Manager neigen zu selbstwertdienlichen Umdeutungen der Realität. Sie rühmen sich ihrer Erfolge und schieben Misserfolge auf die Umwelt. Auf ordnungspolitische Belange – etwa Sicherheit – achten sie tendenziell selektiv, und nehmen die Realität durch einen Filter wahr. Sogar Tatsachen, die für ihre

---

[572] Vgl. nur *Fanto*, Buff L Rev 2001, 249, 260: “well-known CEO tendencies, such as overconfidence and self-aggrandizement”; *Barnard*, U Cin L Rev 2008, 405 nennt das “recurring pathologies of CEOs” (mit Bsp. 414 f.).

[573] *Langevoort* in: Gigerenzer/Engel, Heuristics and Law 2006, 87, 91: “persons who rise in the organization to achieve that kind of power are not ordinary, but rather the survivors in a tournament that presumably weeds out those with inferior cognitive traits”.

[574] *Langevoort* in: Gigerenzer/Engel, Heuristics and Law 2006, 87, 92, der zwar meint, Psychologen seien darüber uneins (ohne Nachw.), andernorts unter Berufung auf “a sizable body of research” aber ausführt, “managers tend to develop biased constructions of the firm’s strategic position […] they will be overconfident”, *Langevoort*, Geo L J 2001, 797, 803.

[575] *Engel*, J Inst Econ 2010, 445, 452: “The effect that is best studied in managers is overoptimism.”; *Langevoort*, U Penn L Rev 1997, 101, 140: “overconfidence [… is] frequently observed in field studies of particular firms.”; *Beckhaus* in: Birk u.a., Unternehmenskauf 2010, 11, 25: „Überschätzung der eigenen Fähigkeiten und Kenntnisse (*overconfidence*) stellt eines der am besten abgesicherten Phänomene der Kognitionspsychologie dar.“

[576] *Busenitz/Barney*, J Bus Vent 1997, 9, 15.

Marktposition unentbehrlich sind – etwa Vorjahresumsätze oder die prozentuale Veränderung der Branchenumsätze – nehmen sie verzerrt wahr.“[577]

In der Praxis wurde solche Selbstüberschätzung etwa für den „Sturz der Giganten“ IBM und General Motors Ende der 1980er verantwortlich gemacht,[578] für die mangelnde Profitabilität von Unternehmenszusammenschlüssen allgemein,[579] für das Scheitern von Disneyland Paris,[580] und für die Klageverfahren gegen Apple, TimeWarner und Polaroid, deren Unternehmensleiter auf Erfolgskurs vermehrt Warnzeichen ignoriert und widrige Informationen unterdrückt hatten.[581] Auch der Kapitalmarkt gilt als Jahrmarkt der Eitelkeiten und Tummelplatz systematischer Selbstüberschätzung von Unternehmensleitern.[582] Leider gibt es, anders als zur experimentellen Forschung, noch keine Synthese der beobachtenden Forschung. In jüngerer Zeit haben zahlreiche Kapitalmarktstudien versucht, Selbstüberschätzung von Unternehmensleitern im Investitions- oder Akquisitionsverhalten der Unternehmen aufzuspüren.[583] Die so gewonnene Evidenz bildete nicht zuletzt einen Anlass für Forderungen zur Reform des US-amerikanischen Gesellschafts- und Übernahmerechts.[584]

---

[577] *Engel*, J Inst Econ 2010, 445, 452 m.w.N. zu jedem Satz des Zitats, und dem wichtigen Nachsatz: “In all of these respects, manager behaviour does not seem to systematically deviate from the behaviour of independent individuals. The managerial setting just gives the respective bias its specific form.”

[578] *Levinson*, Am Psy’st 1994, 428, 431 f.: “That kind of corporate narcissism on the part of General Motors led it to discount the Japanese automakers; in the computer industry, it led IBM to ignore the shift in emphasis from hardware to software.”

[579] So die von *Roll*, J Bus 1986, 197 geprägte Hybris-Hypothese (*hubris hypothesis*); dazu ausf. *Fanto*, Ohio St L J 2001, 1333, 1368 ff. m.w.N.; *Homberg/Osterloh*, JfB 2010, 269, 274 ff. m.w.N.; vgl. auch *Hall*, Manage L 2007, 93, 99; *DellaVigna*, J Econ Lit 2009, 315, 342: “Managerial overconfidence provides one explanation for the underperformance of companies undertaking mergers.”; zu einer anderen oben bei Fn. 555.

[580] *Hall*, Manage L 2007, 93, 96 f.

[581] *Langevoort*, U Penn L Rev 1997, 101, 146 f.: “In most of these cases, a highly successful organization […] was sued for not disclosing some bits of adverse information later found in the company’s files.”

[582] *Black*, Stan L Rev 1989, 597, 624; umgekehrt vermutet *Langevoort* in: Gigerenzer/Engel, Heuristics and Law 2006, 87, 93 Fn. 2, dass kapitalmarkt- und wettbewerbsferne Unternehmen – etwa im Versorgungssektor – Selbstüberschätzung weniger kultivieren.

[583] *Hayward/Hambrick*, Admin Sci Q 1997, 103; *Heaton*, Fin Mgmt 2002, 33; *Simon/Houghton*, Acad Mgmt J 2003, 139; *Malmendier/Tate*, J Fin 2005, 2661; *Malmendier/Tate*, Eur Fin Mgmt 2005, 649; *Lin/Hu/Chen*, Pacific-Basin Fin J 2005, 523; *Brown/Sarma*, J Econ Bus 2007, 358; *Ben-David/Graham/Harvey*, NBER (13711) 2007; *Doukas/Petmezas*, Eur Fin Mgmt 2007, 531; *Trevelyan*, Mgmt Dec 2008, 986; *Malmendier/Tate*, J Fin Econ 2008, 20; *Wang/Zhang/Yu*, Front Bus Res China 2009, 453; *Martin*, Corporate Decisions 2008; *Chatterjee/Hambrick*, Admin Sci Q 2007, 351 messen den Narzissmus des CEO gar an seiner Präsenz in öffentlichen Unternehmensdokumenten.

[584] *Fanto*, Buff L Rev 2001, 249, 285 ff.; *Paredes*, Fla St U L Rev 2005, 673, 736 ff.

Die empirische Forschung lässt also den Schluss zu, dass Selbstüberschätzung bei Unternehmensleitern durchaus weit verbreitet ist, und führt zu der Frage, wie sich die kollegiale Tätigkeit auf das Maß dieser Selbstüberschätzung auswirkt.

2. Gruppenforschung zur Selbstüberschätzung

Leider existieren nur wenige empirische Studien, die untersuchen, wie sich Gruppenarbeit auf die Selbstüberschätzung der Gruppenmitglieder auswirkt. Interessant ist zunächst eine beobachtende Studie dreier deutscher Autoren.[585] Im Gegensatz zu früheren Kapitalmarktstudien, die im US-amerikanischen Kontext stets nur die Urteilsverzerrungen einzelner Unternehmensleiter betrachteten – zumeist derjenigen mit überragender Entscheidungsgewalt (CEO oder CFO) –, widmete sich die deutsche Studie der Selbstüberschätzung *aller* Vorstandsmitglieder. Diese bemaßen sie daran, ob ein Vorstandsmitglied einen positiven Jahressaldo von Ankäufen nach § 15a I 1 WpHG (*directors' dealings*) verzeichnete.[586] Denn nach der Portfoliotheorie sollten sich risikoneutrale Vorstandsmitglieder diversifizieren und ihre Bindung an das Unternehmen (*firm specific risk*) nicht noch über Anstellungsvertrag und Optionsvergütung hinaus erhöhen; wer stattdessen Aktien des eigenen Unternehmens kaufe, müsse sie für unterbewertet halten, also erwarten, eine höhere als die marktübliche Rendite erwirtschaften zu können.[587] Die Studie stellte fest, dass diese Art der Selbstüberschätzung (und zwar nicht nur seitens des Vorstandsvorsitzenden, sondern seitens aller Vorstandsmitglieder) tatsächlich das Investitionsverhalten der Nicht-Finanz-Unternehmen im kombinierten deutschen Aktienindex (CDAX) zwischen 2000 und 2006 erklären konnte.[588] Wenngleich Zweifel an der statistischen Auswertung durchaus angebracht sind,[589] illustriert dieser Befund, dass sich die Selbstüberschätzung aller Vorstandsmitglieder durch das Kollegialorgan hindurch auf das Investitionsverhalten des Unternehmens auswirken kann. Mehr als eine Illustration liegt darin allerdings nicht, denn die Studie kann schon nicht ausschließen, dass sich diese

---

[585] *Glaser/Schäfers/Weber*, SSRN 2008.

[586] *Glaser/Schäfers/Weber*, SSRN 2008, 9 f.

[587] *Glaser/Schäfers/Weber*, SSRN 2008, 6 bei Fn. 4 (unter Berufung auf *Malmendier/Tate*, J Fin 2005, 2661) und 19 ff. zu anderen möglichen Motiven (dazu auch *Homberg/Osterloh*, JfB 2010, 269, 281).

[588] *Glaser/Schäfers/Weber*, SSRN 2008, 13: "The effect is mainly driven by CEO optimism in the case of capital expenditures. However, optimism of other managers has also explanatory power."

[589] Die Autoren stützen ihr Ergebnis auf Regressionskoeffizienten, die bei über 2.600 Beobachtungen nur in etwa der Hälfte der Spezifikationen auf einem 10-Prozent-Niveau signifikant werden.

Selbstüberschätzung ohne Kollegialorgan vielleicht noch stärker niedergeschlagen hätte; dazu bedürfte es einer gezielten Untersuchung der betreffenden Ursachenzusammenhänge.

Näheren Aufschluss über die Ursachenzusammenhänge könnten abermals experimentelle Studien beisteuern. Der einschlägige Forschungsbericht stellt fest, dass Gruppen „allgemein selbstsicherer als Individuen" seien, dass aber unklar sei, ob das auf Selbstüberschätzung beruhe.[590] Denn von drei betrachteten Primärstudien belege die früheste zwar, dass Gruppenmitglieder weniger zu Selbstüberschätzung neigen als Individuen; eine spätere hingegen finde genau den umgekehrten Effekt, und die jüngste belege Selbstüberschätzung auf ähnlichem Niveau bei Gruppen wie bei Individuen.[591] Eine weitere Studie zeigt, dass schon die Aussicht darauf, Ergebnisse in einer Gruppe begründen zu müssen, die Selbstsicherheit reduzieren und dadurch zu geringerer Selbstüberschätzung führen kann.[592] Andererseits sind auch Fälle belegt, in denen falsch antwortende Gruppen sogar höhere Selbstsicherheit angeben als richtig antwortende Individuen, und in denen die Mitglieder solcher Gruppen bei späterer Alleinarbeit die Selbstüberschätzung der Gruppe übernehmen.[593]

Wenngleich die genannte Literatur also auf den ersten Blick keine allzu klare Aussage über die Wirkung von Gruppen auf das Maß der Selbstüberschätzung erlaubt, konnte eine darauf aufbauende Computersimulation[594] Einflussfaktoren dafür identifizieren, wie sich Gruppenarbeit auf die Selbstüberschätzung auswirkt. Zu diesen Faktoren gehören insbesondere die anfängliche Verteilung von individuellen Präferenzen sowie die Aggregationsregel, nach der die Gruppenentscheidung gebildet wird.[595] Auch die Gruppengröße hat einen Einfluss, allerdings nicht auf die Richtung des Effekts, sondern nur auf dessen Größe.[596] Schließlich ist auch die Aufgabenart ein wichtiger – wenn nicht sogar entscheidender – Einflussfaktor.

---

[590] *Kerr/MacCoun/Kramer*, Psy Rev 1996, 687, 692 (Tabelle 1 a.E., r.Sp.); alternativ könnten Gruppen schlicht besser entscheiden und deshalb zu recht selbstsicherer sein, *Sniezek*, Org Behav Hum Dec Proc 1992, 124, 137.

[591] *Kerr/MacCoun/Kramer*, Psy Rev 1996, 687, 692 m.w.N. (Tabelle 1 a.E., l.Sp.); die dort als „unveröffentlicht" zit. Studie „Dunning & Ross (1992)" ließ sich durch Nachfrage bei den Studienautoren und einiger Recherche dennoch auffinden: *Dunning/Ross*, BEDR 1990, 90-02.

[592] *Arkes/Christensen u.a.*, Org Behav Hum Dec Proc 1987, 133, 139 f.; der Effekt scheint aber eher gering.

[593] *Puncochar/Fox*, J Edu Psy 2004, 582, 590 mit dem provokativen Etikett "'two heads' are worse than one effect".

[594] Näher zu Methodik und Konfiguration dieser Simulation *Kerr/MacCoun/Kramer*, Psy Rev 1996, 687, 693 ff.

[595] *Kerr/MacCoun/Kramer*, Psy Rev 1996, 687, 702, 713.

[596] *Kerr/MacCoun/Kramer*, Psy Rev 1996, 687, 702: "direction of the difference […] is generally unaffected by group size, but this difference tends to get larger as group size increases."

Die Bedeutung der Aufgabenart war Gegenstand einer Studie von 1997.[597] Die Studie folgte der klassischen Differenzierung zwischen Intelligenz- und Beurteilungsaufgaben (siehe oben bei Fn. 431) und unterteilte letztere wiederum danach, ob sie nur persönliche Meinungen abfragen und gar keine richtige Lösung haben, oder ob sie schlicht zu komplex sind, um eine existierende richtige Lösung zu demonstrieren.[598] In der Studie mussten Teilnehmer eine Mischung von 20 Auswahlfragen aus diesen drei Kategorien beantworten und jeweils angeben, wie sicher sie sich ihrer Antwort waren. Anschließend wurden die Teilnehmer in Experimentalbedingungen aufgeteilt und erhielten noch einmal dieselbe Aufgabe. Allerdings musste nur die Kontrollgruppe allein arbeiten; in den beiden anderen Bedingungen sollten die Teilnehmer in Zweier- oder Fünfergruppen arbeiten, die frei kommunizieren konnten.[599] Die Ergebnisse belegen zunächst, dass Individuen nur bei Intelligenzaufgaben – nicht aber bei objektiv lösbaren Beurteilungsaufgaben – deutlich selbstsicherer sind, wenn sie die richtige Lösung tatsächlich gefunden haben.[600] Zugleich haben die selbstsichersten Individuen den größten Einfluss auf die spätere Gruppenentscheidung, und zwar ganz unabhängig davon, ob die von ihnen vorgeschlagene Lösung tatsächlich richtig ist – und umso mehr, je deutlicher die Aufgabe Intelligenz statt Urteilsfähigkeit verlangt.[601] Das führte dazu, dass die Selbstüberschätzung der Gruppe bei Intelligenzaufgaben geringer, bei Urteilsaufgaben hingegen höher ausfiel als die des Einzelnen, und zwar jeweils umso deutlicher für die größere Gruppe.[602]

Die Übertragbarkeit dieser Erkenntnis in den Kontext von Kollegialorganen könnte freilich dadurch beschränkt sein, dass Vorstandsmitglieder eine höhere und homogenere Veranlagung zur Selbstüberschätzung haben als der Bevölkerungsdurchschnitt (oben C.V.1.), daher mag der übermäßige Einfluss des selbstsichersten Individuums innerhalb von Vorstandsgruppen

[597] *Zarnoth/Sniezek*, J Exp Soc Psy 1997, 345.

[598] *Zarnoth/Sniezek*, J Exp Soc Psy 1997, 345, 346: "applying the intellective-judgmental continuum of decision tasks"; auch *Kerr/MacCoun/Kramer*, Psy Rev 1996, 687, 697 legen diese Differenzierung zugrunde, indem sie Intelligenzaufgaben als „asymmetrisch" klassifizieren, weil die Minderheit eine Mehrheit überzeugen kann.

[599] *Zarnoth/Sniezek*, J Exp Soc Psy 1997, 345, 350.

[600] *Zarnoth/Sniezek*, J Exp Soc Psy 1997, 345, 356 f.; ähnl. neuerdings *Koriat*, Sci 2012, 360, 362: "in situations in which most participants tend to make the wrong decisions, […] it is the low-confidence individuals who are more likely to be correct, and reliance on the more confident members should lead the group astray."

[601] *Zarnoth/Sniezek*, J Exp Soc Psy 1997, 345, 353 f., 358.

[602] *Zarnoth/Sniezek*, J Exp Soc Psy 1997, 345, 362: "As a result, (*a*) smaller groups may be more overconfident than larger groups on intellective tasks, and (*b*) larger groups may be more overconfident than smaller groups on judgmental tasks."; das könnte wiederum auf den oben (IV.2.) dargestellten Prozessen der Gruppenpolarisierung beruhen.

weniger zum Tragen kommen als in der Experimentalgruppe. Darüber lässt sich allerdings einstweilen nur spekulieren.

VI. Zusammenfassung: Was hält das Kollegialprinzip?

Der Stand der empirischen Forschung lässt sich wie folgt zusammenfassen:

Motivation durch Partizipation? Wie Gruppenzugehörigkeit die Einsatzbereitschaft des Einzelnen beeinflusst, hängt von mehreren Faktoren ab. Belegt sind sowohl Steigerungen als auch Minderungen der Leistungsbereitschaft. Steigerungen treten insbesondere in einfachen Aufgaben sowie dann auf, wenn schwächere Gruppenmitglieder ihren Beitrag für unverzichtbar und es gleichzeitig für möglich halten, durch größeres Engagement ihren Fähigkeitsunterschied auszugleichen. Das ist in Entscheidungsaufgaben und Beschlüssen eher selten der Fall. Zudem sind Motivationssteigerungen eher bei körperlichen als bei geistigen Anstrengungen belegt. Demgegenüber sind gerade bei kognitiv anspruchsvollen oder komplexen Aufgaben sowie in Beschlusssituationen, in denen der einzelne Beitrag nicht identifizierbar oder nicht ausschlaggebend ist, primär Reduktionen der Leistungsbereitschaft in den Formen des sozialen Trittbrettfahrens und sozialen Faulenzens belegt.

Synergien in der Entscheidungsfindung? Gruppen neigen dazu, vor allem geteilte Informationen zu diskutieren. Das führt zu einer mangelnden Sensibilität für die verteilten Informationen einzelner Gruppenmitglieder. Wenn vor allem verteilte Informationen für die beste Entscheidungsalternative sprechen (sog. verborgene Informationsprofile) steht die Gruppendynamik damit einer Entscheidung für die beste Alternative im Weg. Ist diese überhaupt *nur* anhand der verteilten Informationen der Gruppe zu erkennen, also keinem Gruppenmitglied von vornherein einsichtig, gelingt es der Gruppe nur in ganz seltenen Fällen, die beste Alternative zu erkennen. Dabei handelt es sich keineswegs um Laborartefakte, doch könnte die praktische Relevanz der durch verteilte Informationen hervorgerufenen Probleme dadurch vermindert werden, dass Gruppen mit der Zeit ein arbeitsteiliges Gedächtnis entwickeln, das verteilte Informationen strukturiert und dadurch leichter verfügbar macht. Allerdings gibt es Anhaltspunkte dafür, dass größere Vertrautheit der Gruppe ihr arbeitsteiliges Gedächtnis eher verschlechtert. Ob arbeitsteilige Gedächtnisse also tatsächlich verborgenen Informationsprofilen vorbeugen, ist derzeit allenfalls eine Vermutung, die weiterer Forschung bedarf.

Entschleunigung der Informationssuche? Gruppen benötigen oft länger für eine Entscheidung, können bei der Problemlösung aber auch schneller sein als Einzelne. Betrachtet man als Maß der Geschwindigkeit hingegen nicht die auf der Uhr abgelaufene Zeit, sondern die Menge der zusätzlich berücksichtigten Informationen, deuten einschlägige Experimente darauf

hin, dass Gruppen weder langsamer noch schneller entscheiden als Einzelne. Zumindest wenn die längere Informationssuche mit Kosten verbunden ist, sammeln Gruppen genau so viele Informationen wie Einzelne, bevor sie eine Entscheidung treffen. Da diese Schlussfolgerung auf sehr wenigen Studien beruht, bietet sie keine allzu robuste Grundlage für Verallgemeinerungen, ruft aber zumindest Zweifel daran wach, dass Gruppen konservativer und mit größerem Beharrungsvermögen entscheiden als Einzelne.

Gegenseitige Überwachung und Moderation? Gruppen neigen ganz allgemein dazu, die moderaten Einstellungen, die ihre Mitglieder teilen, in Richtung des Extrems zu polarisieren. Das liegt daran, dass geteilte Grundeinstellungen in der Gruppenberatung mit größerer Wahrscheinlichkeit zum Ausdruck kommen als abweichende Tendenzen, und dass die Gruppenmitglieder um soziale Zugehörigkeit wetteifern. Dagegen bedarf es keiner elitären oder hochkohäsiven Gruppe, wie früher von der Theorie des Gruppendenkens angenommen wurde. Gegenseitige Kontrolle wird durch diese Festigung und Verstärkung der Grundeinstellungen ebenso erschwert wie gegenseitige Moderation. Im Gegenteil belegen Studien, dass Gruppenberatung etwa die Neigung zur Einsatzeskalation verstärkt, die ihrerseits gravierende wirtschaftliche Fehlentscheidungen verursachen kann. Während der polarisierende Effekt von Gruppen mittlerweile robust belegt und durch eine Metastudie quantifiziert ist, sind Studien zur Verstärkung der Einsatzeskalation bislang noch rar, insoweit ist also eine vorsichtige Interpretation der Befunde geboten.

Verringerung der Selbstüberschätzung? Unternehmensleiter unterliegen in erheblichem Umfang systematischer Selbstüberschätzung. Ob die Gruppe daran etwas ändert, lässt sich nur mit Blick auf die zu beurteilende Aufgabenart feststellen. In einer einschlägigen Studie steigerten Gruppenmitglieder ihre Selbstüberschätzung vor allem dann, wenn sie in einer größeren Gruppe mit Beurteilungsaufgaben konfrontiert waren. Das entspricht am ehesten der Situation von Leitungsorganen, die oft in größerer Runde Fragen beurteilen müssen, zu denen kein Mitglied eine richtige Lösung kennt. Stattdessen bringt jeder seine persönliche Einschätzung und Selbstsicherheit in die Runde, und durch die Kommunikation ihrer Teilnehmer gelangt die Gruppe zu einer Entscheidung. In diesem Umfeld legt das geschilderte Experiment nahe, dass eine Gruppe stärker unter Selbstüberschätzung leidet als ihre Mitglieder. Letztlich sind endgültige Aussagen zu diesem Fragenkomplex kaum möglich, denn das geschilderte Experiment – wiewohl einschlägig und sehr anschaulich – behebt nicht den festgestellten Mangel an widerspruchsfreier Evidenz. Dennoch sollte es Zweifel daran wecken, ob das Kollegialprinzip der ihm zugedachten Aufgabe gerecht werden kann, die Selbstüberschätzung der Organmitglieder einzuhegen.

D. Dogmatische und rechtspolitische Schlussfolgerungen

Nach der systematischen Sichtung des Stands der empirischen Forschung sind die gewonnenen Erkenntnisse nunmehr an die rechtliche Ausgangsfrage rückzubinden. Dabei werde ich zunächst auf den Versuch eingehen, die verschiedenen Verhaltenseffekte miteinander zu kombinieren (I.). Im Anschluss sind rechtsdogmatische (II.) und rechtspolitische (III.) Schlussfolgerungen zu erörtern.

I. Zusammenschau verschiedener empirischer Erkenntnisse

Die Darstellung einzelner Verhaltenseffekte darf nicht darüber hinwegtäuschen, dass solche Effekte in der Realität nicht isoliert auftreten. Vielmehr stehen sie in mannigfacher Wechselbeziehung zueinander.[603] Auf einige Querverbindungen wurde bereits hingewiesen,[604] aber über die meisten Wechselwirkungen gibt die empirische Forschung kaum Aufschluss. Zahlreiche psychologische Konzepte sind erprobt, aber übergreifende Verhaltenstheorien fehlen zumeist.[605] Womöglich ist die Zeit dafür noch nicht reif, womöglich kann es sie auch gar nicht geben.[606]

Der am Zusammenspiel solcher Effekte interessierte Sozialwissenschaftler oder Jurist ist damit letztlich auf Vermutungen angewiesen, wie die dargestellten Effekte in komplexeren Situationen zusammentreffen. Ein interessantes Szenario, das Selbstüberschätzung, Gruppenpolarisierung und Einsatzeskalation zueinander in Beziehung setzt, stammt von Donald Langevoort. Danach sammeln gerade erfolgreiche Unternehmensleiter ein enormes Selbstwertgefühl an,[607] weshalb

> „Unternehmensleiter, die eine riskante Maßnahme in Angriff nehmen, anfänglich eine überhöhte Vorstellung von ihrer Fähigkeit haben werden, das Risiko zu beherrschen – wenn sie es überhaupt wahrnehmen. Ermutigungen in der Anfangsphase (d.h. fehlende Andeutungen möglichen Unheils) bestärken diese Illusion. Anzeichen für Gefahr, die mitten in der Ausführung zutage treten, werden die Unternehmensleiter ungern zur

[603] *Langevoort*, U Penn L Rev 1997, 101, 146: "Although each of the foregoing behavioral biases is interesting by itself, the practical implications follow from their interplay."

[604] Oben Fn. 548, 579 a.E., 602 a.E. und bei Fn. 468, 550.

[605] *Lempert*, J Emp L Stud 2010, 907, 911: "Since the world is complex and outcomes often have multiple causes, considerable research may be necessary to develop a theory complete enough to provide a reliable guide to policy."

[606] Menschliches Verhalten ist eine komplexe stochastische Interaktion biologischer und sozialer Systeme. Physik und Chaosforschung können (noch?) nicht einmal *deterministische* Naturgesetze in komplexen Interaktionen vollständig beschreiben, vgl. *Struck*, JuS 1993, 992; *Gleick*, Chaos 1994; *Springer*, SdW 2012, 9/48.

[607] *Langevoort*, U Penn L Rev 1997, 101, 147; zust. *Paredes*, Fla St U L Rev 2005, 673, 675 f.

Kenntnis nehmen, oder sie verleugnen deren Bedeutsamkeit. Sie haben sich für die Maßnahme eingesetzt, und eine Anerkennung der Gefahren würde nachträglich die Angemessenheit ihres ursprünglichen Beschlusses in Frage stellen. Dieses Hochschaukeln zwischen Selbstüberschätzung und Einsatzeskalation wird das Unternehmen allmählich in viel größere Schwierigkeiten bringen.“[608]

Nach *Langevoort* habe just diese Wechselwirkung verschiedener Urteilsverzerrungen maßgeblich zur Enron-Insolvenz beigetragen.[609] Bewusste Irreführung und Täuschung seien dabei stets erst der letzte Schritt einer langen Entwicklung, in deren Verlauf die Unternehmensleiter sich immer weniger von ihrem zunehmend optimistischen Gebahren distanzieren können.[610] Eine ähnliche Dynamik beschreibt Harry Levinson:

„Je höher man in einer Organisation aufsteigt, desto mehr Selbstsicherheit wird man wahrscheinlich entwickeln, die eigenen Rollen auch kompetent ausfüllen zu können. Gleichzeitig gewärtigt man umso weniger Überwachung je höher man steigt. Das Zusammenspiel dieser Faktoren führt oft zu einer narzisstischen Übersteigerung, die sich in Selbstüberschätzung und Anspruchsdenken äußert. Das wiederum führt zur Leugnung von Realitäten, die das übersteigerte Selbstbild bedrohen, und zur Verachtung anderer Personen und Organisationen. Es führt auch zu geringerer Toleranz für Abweichungen vom bisher erfolgreichen Schema.“[611]

Solche komplexen Verhaltensdynamiken könnten klare Aussagen über die Wirkungen des Kollegialprinzips erheblich erschweren. Allerdings ist das genaue dynamische Zusammenspiel verschiedener Effekte bislang schlicht nicht geklärt. Die empirische Forschung kann sich der Frage zwar dadurch *nähern*, dass sie in kontrollierter Umgebung die Interaktionen verschiedener Verhaltenseffekte erkundet und daraus Vorhersagen ableitet, die sich an beobachtenden Daten überprüfen lassen. Juristen hingegen können allenfalls plausible Spekulationen anstellen, sind aber „nicht in der Position, unter widerstreitenden empirischen oder deskriptiven Behauptungen zu entscheiden“.[612] Schließlich hängt „überall so viel von Dingen ab, die sich

---

[608] *Langevoort* in: Gigerenzer/Engel, Heuristics and Law 2006, 87, 97 f.; vgl. auch *Langevoort*, U Penn L Rev 1997, 101, 147; im Original verwenden beide Texte die Metapher *optimism-commitment whipsaw* (wörtlich „Schrotsäge von Optimismus und Einsatz“); zust. *Paredes*, Fla St U L Rev 2005, 673, 691 f.; *Barnard*, U Cin L Rev 2008, 405, 423.

[609] *Langevoort* in: Gigerenzer/Engel, Heuristics and Law 2006, 87, 97 f.; ebenso *Barnard*, U Cin L Rev 2008, 405, 414 m.w. Bsp.

[610] *Langevoort*, U Penn L Rev 1997, 101, 147: “managers were committed to their publicly expressed optimism, from which they could not easily step away, even as the signs of trouble became palpable. Only at that late stage was there a truly deliberate form of dissembling.”; ebenso *Paredes*, Fla St U L Rev 2005, 673, 691: “the CEO otherwise would have to backtrack publicly from a particular decision or overall business strategy.”

[611] *Levinson*, Am Psy’st 1994, 428, 432.

[612] *Langevoort* in: Gigerenzer/Engel, Heuristics and Law 2006, 87, 92 Fn. 1; nach *Garrn*, Rationalität 1986, 115 f. seien solche Entscheidungen „nur auf der Basis eines als selbstverständlich vorausgesetzten empirischen Wissens“ möglich, dessen Bereich „schmal sein kann“.

kaum in Regeln bringen lassen, dass eine abstrakte Erörterung wenig fruchtbar scheint."[613] Deshalb blende ich jene komplexen und bislang kaum erforschten Wechselwirkungen aus und verweile stattdessen beim einigermaßen gesicherten Forschungsstand, der auch ohne weitergehende Spekulationen reichhaltige Erkenntnis bietet.

II. Der rechtsdogmatische Wert der vorliegenden Untersuchung

Empirische Erkenntnis hilft Juristen bei der Durchmusterung ihres Argumentationsarsenals.[614] Die für das Kollegialprinzip angeführten Argumente (oben A.IV.) haben sich als unterschiedlich belastbar herausgestellt (oben C.VI.), was natürlich ihren Stellenwert im normativen Diskurs beeinflussen muss. Wenn die bisherige Dogmatik zum Kollegialprinzip von unzutreffenden empirischen Annahmen ausging, müssen die anhand solcher Annahmen begründeten Rechtsinstitute kritisch hinterfragt und mit einer empirisch besser fundierten Dogmatik rationalisiert werden. Die folgenden Abschnitte bewerten einige der oben aufgeworfenen Zweifel an den empirischen Annahmen des Kollegialprinzips, legen aber auch die Grenzen dieser Zweifel dar. Daraus ergeben sich zwangsläufig weitere Untersuchungsrichtungen und neue Forschungsfragen.

1. Kollegialorgan als Motivationsquelle oder Hängematte

Die in vielen Kollegialorganen behandelten Sachfragen sind „von einer besonderen Komplexität und Unstrukturiertheit sowie risikobehaftet",[615] was empirisch eher zu sozialer Hemmung als zu sozialer Erleichterung führt. Zudem erschwert die kollegiale Beschlussfassung die Identifizierbarkeit einzelner Beiträge, denn

> „Vorstandsmaßnahmen […] gelten als gesamtheitliche Leistung aller Mitglieder, an der der Wert oder Unwert des Beitrages der einzelnen Mitglieder kaum anteilig feststellbar ist. Daraus folgt die Schwierigkeit, die Haftung der einzelnen Mitglieder anteilig zu berechnen."[616]

was zwar aus anderen Gründen durchaus erwünscht sein mag,[617] aber auch die Wahrscheinlichkeit sozialen Faulenzens erhöht. Das gilt in der traditio-

---

[613] *Körner*, ZöR 1937, 55, 56.

[614] So schon *Schweizer* in: GfK e.V., Jahrbuch 1976, 386, 401: „dass bestimmte denkbare Rechtsauslegungen abgelehnt werden"; ähnl. *Rachlinski*, Cornell L Rev 2011, 901, 919: "Empirical studies cannot always answer the ultimate question, but they can rule out certain arguments."; vgl. aber *Tontrup* in: Engel/Schön, Proprium 2007, 192, 200 zur juristischen Immunisierung gegen den „Einwand, eine dogmatische Konstruktion sei empirisch nicht mehr zu halten".

[615] *Grundei/von Werder*, AG 2005, 825, 826.

[616] *Dose*, Vorstandsmitglieder 1975, 115.

nell funktionalen Vorstandsgliederung vielleicht noch mehr als in der heute üblichen divisionalen Spartengliederung,[618] aber auch in jener neueren Form der Vorstandsorganisation bedeutet das Kollegialprinzip „kollegiale" statt nur „paralleler" Unternehmensleitung.[619] Daher verbietet sich die sichtbare Abgrenzung des Beitrags einzelner Vorstandsmitglieder zu Führungsentscheidungen – was als eine der wichtigsten Maßnahmen gilt, um sozialem Faulenzen vorzubeugen.[620] Zudem sind Kollegialorgane zumeist Treuhänder fremden Vermögens, so dass sie die Folgen ihres Handelns nur begrenzt tragen müssen. Das könnte Prozesse der sozialen Hemmung noch zusätzlich begünstigen.[621]

Dem wird entgegengehalten, die Arbeit im Kollegium sei zumindest für professionelle Unternehmensleiter die wichtigste Verpflichtung ihres Berufslebens, deshalb sei die Trittbrettfahrermentalität deutlich seltener als eine Wettbewerbsmentalität – das Bemühen, andere auszustechen und, wenn schon nicht die Gruppe auf die eigene Seite zu ziehen, so doch zumindest gut informiert und intelligent zu erscheinen.[622] Allerdings ist fraglich, ob diese Eindruckssteuerung (*impression management*) langfristig aufrechterhalten wird, wenn sie lediglich durch das Erfolgserlebnis innerhalb des Organs belohnt und nicht nach außen sichtbar wird. Wichtiger er-

---

[617] Vgl. schon *von Jhering*, Zweck 1904, 402: „Niemand kann ein einzelnes Mitglied mit Sicherheit [für einen Kollegialbeschluss] verantwortlich machen, und diese Ungewissheit, dieser Schleier" sichere gerade die Unabhängigkeit kollegial entscheidender Richter.

[618] Zur Ressortpraxis *Martens* in: Goerdeler u.a., FS Fleck 1988, 191, 192; *Bleicher*, zfbf 1988, 930, 940: „weitgehend übertriebene Ressortierung"; instruktiv zur Praxis auch *von Hein*, ZHR 2002, 464, 482 ff.

[619] So *von Hein*, ZHR 2002, 464, 485; *Dauner-Lieb* in: Crezelius u.a., FS Röhricht 2005, 83, 99; tatsächlich allerdings ist nach *Rieger* in: Lutter u.a., FS Peltzer 2001, 339, 347 „die Sachnähe und Fachkompetenz des Spartenvorstands in der Regel so stark, dass die anderen […] keine besondere Veranlassung verspüren, die Beschlussanträge des Spartenkollegen zu hinterfragen. Das kann zu gravierenden Defiziten in der Wahrnehmung der gemeinschaftlichen Verantwortung des Vorstands führen."; ebenso *Bleicher/Leberl/Paul*, Unternehmensverfassung 1989, 272 f.: „der *Ressortspezialist* […] versteht von ‚seinem Geschäft' […] weit mehr als seine Kollegen, und ist nicht selten mehr an der *Durchsetzung von Bereichsinteressen* – denn das verschafft ihm Ansehen und Unterstützung bei ‚seiner Mannschaft' – als am Konsens im Hinblick auf *übergeordnete Gesamtinteressen* der Unternehmung interessiert."; prägnant *Bernhardt/Witt*, ZfB 1999, 825, 828: „das Vorstandssystem denaturiert zu einer Versammlung nebeneinander regierender CEO's".

[620] *Karau/Williams*, J Pers Soc Psy 1993, 681, 700: "monitoring individual performance or making such performance identifiable"; vgl. auch *Schüller*, Vorstandsvergütung 2002, 97: „Die Zuordnung einer bestimmten Leistung zu einer bestimmten Person innerhalb des Kollektivorgans ist [.] wichtig für die Motivationswirkung".

[621] *Karau/Williams*, J Pers Soc Psy 1993, 681, 700: "individuals are more likely to engage in social loafing […] when working on tasks that are perceived as low in meaningfulness or personal involvement".

[622] *Blinder*, Int Fin 2009, 171, 174.

scheint demgegenüber das Argument, Unternehmensleiter seien intrinsisch motiviert, „da ihr Selbstwertgefühl von hohem Arbeitseinsatz und Erfolg abhängt."[623] Allerdings dürften sich die meisten empirischen Belege für dieses Argument auf den US-amerikanischen CEO oder prominente Vorstandsmitglieder beziehen, während für die weniger sichtbare Vorstandsriege durchaus fraglich ist, ob ihre intrinsische Motivation die nachteilige Gruppendynamik im Kollegialorgan überwinden kann. Ebenso ließe sich annehmen, dass die Gruppendynamik weniger starke Einsatzsteigerungen bewirkt als sie zugleich Einsatzminderungen verursacht. Angesichts dieser empirisch fundierten Zweifel kann die motivierende Wirkung der Gruppenentscheidung nicht als tragfähige dogmatische Rechtfertigung für das Kollegialprinzip herhalten.

2. Informationsverarbeitung durch Kollegialorgane

Nach dem Stand der empirischen Erkenntnis ist es zweifelhaft, ob Kollegialorgane bei der Informationssuche konservativer und weniger vorschnell handeln als einzelne Entscheidungsträger. Wenngleich es voreilig erscheint, daraus eine *Widerlegung* der gesetzgeberischen Intuition zu konstruieren,[624] verschiebt diese Erkenntnis doch das Hauptaugenmerk von der Geschwindigkeit auf den Inhalt und die Fundierungsprozesse der Kollegialentscheidung.

Welchen Beitrag das Kollegialprinzip zur besseren Fundierung von Entscheidungen treffen kann, bestimmt sich maßgeblich im kritischen Grenzfall der verborgenen Informationsprofile. Die experimentelle Forschung zu verborgenen Informationsprofilen sieht sich – wie jede experimentelle Forschung mit Praxisbezug – Einwänden hinsichtlich ihrer äußeren Gültigkeit ausgesetzt. So wird ihr etwa entgegengehalten, dass sie mindestens sechs Umstände realer Gruppenentscheidungen nicht reflektiere:[625] erstens dass Gruppenmitglieder nicht alle Informationen gleich wertschätzen sondern bestimmte Informationsarten womöglich bevorzugen, zweitens dass Gruppenmitglieder untereinander Interessenkonflikten und Zieldivergenzen unterliegen können, drittens dass Gruppenmitglieder gezielt beeinflussen, *wie* sie Informationen preisgeben und mit welchem Unterton, viertens dass geteilte Informationen oft einfach relevanter sind als die verteilten, fünftens

---

[623] *Arnold*, Steuerung 2007, 52, 131, je m.w.N.; ebenso das Befragungsergebnis (dazu ausf. § 3 B.IV.2.b) von *Pepper/Gore/Crossman*, Hum Res Mgmt J 2013, 36, 41: "most executives are driven by a sense of achievement […] summarised in the words of one participant as 'winning' […] Chief executives, competitive by nature, want to know how they are doing relative to their peers."

[624] So *Bainbridge*, Vand L Rev 2002, 1, 14: "[…] tends to disprove the common intuition that it takes groups longer to make decisions."

[625] *Wittenbaum/Hollingshead/Botero*, Comm Mono 2004, 286, 298 ff.

dass Informationen ganz anders verteilt sein können als in verborgenen Profilen, und sechstens dass Gruppenmitglieder gezielt beeinflussen, *wem* sie Informationen preisgeben. Interessanterweise geht diese Kritik gar nicht davon aus, dass die Forschung mit realitätsnäheren Annahmen „bessere" Entscheidungen beobachten würde[626] – und in der Tat ist zweifelhaft, ob etwa gruppeninterne Interessenkonflikte anstelle des gemeinsamen Interesses an der bestmöglichen Gruppenentscheidung zu einer höheren Entscheidungsqualität führen.[627] So wird dem Kollegialprinzip in der Praxis bescheinigt, es verursache „Reibungsverluste durch (Fehl-)Verhaltensweisen von Vorstandsmitgliedern, die sich in Dominanzbestrebungen, Kompetenzstreitigkeiten und Ressortegoismus äußern können."[628] Obwohl die Gruppenforschung solche Interessenkonflikte überwiegend ausblendet,[629] werden ihre Ergebnisse dadurch nicht weniger relevant. Wenn die Gruppe schon in einer abstrakten Idealkonstellation nicht in der Lage ist, verteilte Informationen zusammenzuführen, wie soll es ihr dann in der komplexen Realität besser gelingen?[630]

Sind aber die in der empirischen Forschung dokumentierten gruppendynamischen Prozesse auch in der Realität einer optimalen Entscheidungsfindung abträglich, bleibt freilich die Frage, was daraus folgt. Ohne Kollegialentscheidung gäbe es diese gruppendynamischen Prozesse zwar nicht – ebensowenig gäbe es aber die zur optimalen Entscheidung erforderlichen Informationen. Daher scheint es auf die rhetorische Betonung anzukommen: *Verborgene* Informationsprofile gelten zwar als Hindernis für *gute* Gruppenentscheidungen; zugleich sind verborgene Informations*profile* aber der Grund für *bessere* Gruppenentscheidungen, verglichen mit denen der Gruppenmitglieder.[631] Ohne Gruppenentscheidung gäbe es ja gar kein

---

[626] Nach *Wittenbaum/Hollingshead/Botero*, Comm Mono 2004, 286, 304 gehe es nur um ein "framework for understanding motivated information sharing".

[627] Schon *Haymann* in: Tatarin-Tarnheyden, FS Stammler 1926, 395, 476 betonte, „daß alle Mehrheitsentscheidung nur in der allen Abstimmenden *gemeinsamen Richtung* auf die gleiche Frage ihren Sinn und ihre Schranke empfängt" (Hervorhebung nur hier).

[628] *Fleischer*, NZG 2003, 449, 458; zust. *Beckert*, Personalisierte Leitung 2009, 81 bei Fn. 384; ähnl. *Seibt* in: Lutter/Schmidt, AktG 2010, § 77 Rn. 2; *Seidenfeld*, Cornell L Rev 2002, 486, 535: "the composition, structure, and dynamics of agency work groups often will preclude whatever potential synergies the group might achieve."; *Brodbeck/Kerschreiter u.a.*, Acad Mgmt Rev 2007, 459, 470: "some group members may pursue their own (often hidden) goals in order to maintain or increase their power in general or to implement specific decisions for their own benefit."

[629] *Sniezek*, Org Behav Hum Dec Proc 1992, 124, 125.

[630] Ähnl. *Fiorina/Plott*, Am Polit Sci Rev 1978, 575: "if a given model does not predict well relative to others under a specified set of conditions in the controlled world of the laboratory, why should it receive preferential treatment as an explanation of non-laboratory behavior occurring under similar conditions?"

[631] *Brodbeck/Kerschreiter u.a.*, Acad Mgmt Rev 2007, 459, 462.

Informationsprofil, sondern nur verteilte Informationen. Erst das Zusammenkommen dieser verteilten Informationen ermöglicht die zahlreich empirisch belegten Fälle, in denen Gruppen besser entscheiden als zumindest ihr durchschnittliches, wenn nicht sogar ihr kompetentestes Mitglied.[632] Da das kompetenteste Mitglied in der Praxis selten im Vorhinein identifizierbar oder gar abstrakt bestimmbar ist, führt das Kollegialprinzip die in Frage kommenden Entscheidungsträger zusammen, um die Wahrscheinlichkeit zu erhöhen, dass jedenfalls immer auch der kompetenteste mitbeteiligt ist.[633] Nach dieser Lesart sind abträgliche gruppendynamische Effekte nichts anderes als der Preis, den die Gruppe für die größere *potentielle* Informationsbasis und die *wahrscheinliche* Beteiligung des kompetentesten Entscheidungsträgers zahlt.[634] Mitunter wird sogar versucht, die Aktualisierung dieses Potentials durch Haftungskonsequenzen zu erzwingen:

> „Insbesondere die Teilnahme an der [Vorstands]Beratung und die gegenseitigen Versuche, kontroverse Meinungen [zum Zwecke der] Überzeugung der anderen Mitglieder darzustellen, ist [sic] auch aus gruppendynamischen Aspekten ein wesentlicher Teil der Gesamtgeschäftsführung. Schon aus diesem Grunde würde ein Vorstandsmitglied pflichtwidrig handeln, wenn es einfach den Vorstandssitzungen fern bleibt oder zwar teilnimmt, sich aber an der Geschäftsdiskussion nicht beteiligt."[635]

Indessen belegt die empirische Forschung, dass die Vergrößerung der Informationsbasis besser durch eine klare Spezialisierung der Gruppenmitglieder[636] und durch arbeitsteiliges Zusammenwirken der Gruppe zu erreichen ist, etwa aufgrund der Herausbildung eines transaktiven Gruppengedächtnisses.[637] Im Grenzfall bedeutet eine solche Arbeitsteilung wiederum die Auflösung des Kollegialprinzips. Zu Recht wird daher angemerkt, dass auch in einer monokratischen (direktorialen) Führungsstruktur alle erfor-

---

[632] *Brodbeck/Kerschreiter u.a.*, Acad Mgmt Rev 2007, 459, 470: "a major motivation for using groups as decision-making agents under conditions of distributed knowledge is the hope for higher group decision quality."

[633] *Bainbridge*, Vand L Rev 2002, 1, 26 m.w.N. zur empirischen Identifizierbarkeit des Kompetentesten.

[634] Vgl. *Rieß* in: Jung u.a., FS Müller 2008, 599, 609: „Jede Kollegialentscheidung enthält die *Chance* der Vergrößerung des Argumentationsbestandes" (Hervorhebung nur hier); prägnant *Grundei*, BCCG DiskP 2004, 1/13 f.: das Kollegialprinzip gebe „zunächst nur strukturell die Gelegenheit", viele Perspektiven zu berücksichtigen, „[t]atsächlich kann die Nutzung dieses Potenzials jedoch an verschiedenen gruppenpsychologischen Phänomenen scheitern."

[635] *Dose*, Vorstandsmitglieder 1975, 118.

[636] *Lu/Yuan/McLeod*, Pers Soc Psy Rev 2012, 54, 55 m.w.N.: "designating group members to be experts in some domain of the information reduced, but did not eliminate, the advantage of common over unique information."

[637] Es genügt nicht, sich der Möglichkeit verteilter Informationen nur bewusst zu sein oder die entsprechende Entscheidungsaufgabe nur hinreichend oft zu üben, so *Brodbeck/Kerschreiter u.a.*, Acad Mgmt Rev 2007, 459, 471.

derlichen Informationen von nach- bzw. untergeordneten Stellen beschafft werden könnten, ohne dass es dafür ihrer kollegialen Gleichordnung bedürfte.[638] Gruppendynamische Effekte würden dadurch vermieden. Zwar gewönnen zugleich die individuellen kognitiven Beschränkungen des Entscheidungsträgers überragenden Einfluss auf die Art und Menge der angeforderten Informationen,[639] das passiert allerdings auch in den mit starken Persönlichkeiten besetzten Kollegialgremien, solange der jeweilige Stimmführer nicht tatsächlich kompetenter ist als seine Kollegen.[640]

Abschließend ist daher davon auszugehen, dass Kollegialorgane nicht die einzige zur Zusammenführung (*pooling*) entscheidungsrelevanter Informationen geeignete Entscheidungsstruktur bilden, und dass gruppendynamische Prozesse die Informationsverarbeitung eher beeinträchtigen als erleichtern. Die Informationsverarbeitung in der Gruppe bietet damit ebenfalls keine tragfähige dogmatische Rechtfertigung für das Kollegialprinzip.

3. Kollegialprinzip als Mechanismus des debiasing

Die psychologische Forschung über kognitive Urteilsverzerrungen (*biases*) wird zunehmend auch im deutschen Gesellschafts- und Kapitalmarktrecht rezipiert.[641] Dabei gilt die Kompensation solcher Verzerrungen durch Institutionen (*debiasing*) mitunter als vielversprechender Ansatz, der paternalistische Regulierung mit liberalem Denken vereinigen könnte.[642] Entsprechend wird auch dem Kollegialorgan bisweilen eine solche institutionelle Kompensationswirkung zugeschrieben:

> „Es sind nicht zuletzt [.] auf Schranken der individuellen Informationsverarbeitungskapazität zurückgehende Anomalien, die für die multipersonelle Besetzung von Organisationseinheiten (wie dem Vorstand) sprechen. […] Im günstigsten Fall bestehen komplementäre Stärken und Schwächen der Gruppenmitglieder, so dass Fehler (wie z.B. die genannten Biases) des einen durch ‚Kontrollen' der anderen erkannt und kompensiert werden können."[643]

---

[638] *Groß*, Kollegialprinzip 1999, 204; für die Justiz vgl. *Kissel*, Gerichtsbarkeit 1972, 55: „jeder [Einzel-]Richter wird, wenn er wirklich das Bedürfnis nach einem Gespräch […] mit einem anderen Richter sucht [sic], niemals ohne kollegialen Gesprächspartner bleiben."

[639] So wohl *Grundei*, BCCG DiskP 2004, 1/10: die „Einmannleitung" berücksichtige „tendenziell weniger Wissen", weil nachgeordnete Manager „lediglich für die Bearbeitung ihrer jeweiligen Zuständigkeitsbereiche genutzt" würden.

[640] *Brodbeck/Kerschreiter u.a.*, Acad Mgmt Rev 2007, 459, 471; ebenso *Stasser/Stewart/Wittenbaum*, J Exp Soc Psy 1995, 244, 262: "assigning expert roles on misguided assumptions […] may backfire [… and] easily aggravate the tendency to omit unshared items from discussion."

[641] Z.B. *Fleischer* in: Fuchs u.a., FS Immenga 2004, 575; *Klöhn*, Spekulation 2006, 90 ff.;

[642] *Jolls/Sunstein*, J L Stud 2006, 199; Kritik am *libertarian paternalism* und Nachweise bei *Mitchell*, Nw U L Rev 2005, 1245; aus deutscher Sicht *Eidenmüller*, JZ 2011, 814.

Selten finden sich dagegen konkrete Vermutungen darüber, welche Arten von Urteilsverzerrungen durch das Kollegialprinzip kompensiert werden könnten. Eine Ausnahme bildet die Selbstüberschätzung, der sich die vorangegangenen Ausführungen deshalb ausführlich gewidmet haben. Die empirische Literatur belegt robust, dass Unternehmensleiter besonders stark zur Selbstüberschätzung neigen. Studien zu etwaigen Kompensationswirkungen der Gruppe sind zwar rar, legen aber nahe, dass keine der beiden pauschalen Ansichten verfängt: weder neigen Gruppen allgemein weniger zu Selbstüberschätzung als Individuen,[644] noch erhöhen sie generell die Selbstüberschätzung.[645] Vielmehr ist nach der Art der Aufgabe zu differenzieren, die die Gruppe zu bewältigen hat. Zumindest für diejenige Aufgabenart, die Kollegialorgane regelmäßig gewärtigen, legt die empirische Forschung Zweifel daran nahe, dass die Diskussion in der Gruppe Selbstüberschätzung reduzieren kann. Für eine endgültige Bewertung ist die Evidenz zu dünn, aber auch die Befunde zur Gruppenpolarisierung sprechen eher dagegen, dem Kollegialprinzip in irgendeiner Hinsicht eine moderierende Wirkung beizumessen.[646]

Die normative Bewertung dieser Schlussfolgerung hängt auch von der Frage ab, inwieweit die Selbstüberschätzung von Unternehmensleitern überhaupt als kompensationsbedürftig angesehen wird:

> „Ob bestimmte empirisch feststellbare Phänomene negativ, neutral oder positiv zu bewerten sind, lässt sich nicht allgemein beantworten, und kann in einigen Fällen (wie zum Beispiel dem überzogenen Optimismus) durchaus schwierig zu beantworten sein.“[647]

Einige Stimmen weisen darauf hin, dass Selbstüberschätzung allgemein ein biologisch adaptives Merkmal ist,[648] das gerade auch im Wettbewerb um Führungspositionen einen Selektionsvorteil darstellt.[649] Die Selbstüberschätzung von Unternehmensleitern mag durchaus handfeste Vorteile für

---

[643] *Grundei/von Werder*, AG 2005, 825, 829; ähnl. *Wettich*, Vorstandsorganisation 2008, 300; vgl. auch *Sniezek* (Fn. 566); schon *Haft*, Mich L Rev 1981, 1, 57 folgerte aus der Existenz individueller Urteilsverzerrungen “that corporations are safer committing major business decisions to an intelligent peer group than to any intelligent individual.”

[644] So insb. *Fleischer* (oben Fn. 287) und *Bainbridge* (oben Fn. 289); *Engel*, J Inst Econ 2010, 445, 455: “Groups are not immune, but less liable than individuals to overconfidence”.

[645] So etwa *Langevoort* (oben Fn. 290).

[646] Andere, vielleicht wirksamere, Maßnahmen nennen *Heath/Larrick/Klayman*, Res Org Behav 1998, 1, 20 ff.

[647] *Englerth* in: Engel u.a., Recht und Verhalten 2007, 60, 101.

[648] *Heath/Jourden*, Org Behav Hum Dec Proc 1997, 103 m.w.N.

[649] *Engel*, J Inst Econ 2010, 445, 459 f. verweist auf Unternehmen, die gezielt überoptimistische Vertriebsmitarbeiter einstellen; ausf. zur Deutung als Selektionsvorteil *Langevoort*, U Penn L Rev 1997, 101, 152 ff.; *Paredes*, Fla St U L Rev 2005, 673, 684 f.; *Langevoort* in: Gigerenzer/Engel, Heuristics and Law 2006, 87, 92 ff.; ähnl. *Levinson*, Am Psy’st 1994, 428, 430: “winners in this socioeconomic combat”.

ihre Unternehmen haben, etwa indem sie zu „zu höherer Motivation, größerer Ausdauer und zu mehr Risikobereitschaft“ beiträgt,[650] Tatkraft und Entschlussfreude befördert oder eine optimistische Unternehmenskultur schafft.[651] Sowohl die Motivation der Mitarbeiter als auch das Bestehen gegen Wettbewerber im Markt könnte maßgeblich davon abhängen, welche Überzeugung und Risikobereitschaft die Unternehmensführung ausstrahlt.[652] Ganz ähnlich stellt die beobachtende Teamforschung fest, dass höhere Selbstsicherheit der Gruppe (*efficacy* bzw. *potency*) mit größerem Erfolg einhergeht.[653] Aus rechtsökonomischer Perspektive schließlich kann die Selbstüberschätzung von Unternehmensleitern womöglich Agenturkosten (*agency costs*) vermeiden helfen:[654] Zum einen schafft sie Risikofreude, die wohldiversifizierte Aktionäre nach der Portfoliotheorie zwar bevorzugen, Unternehmensleiter dagegen scheuen, weil sie ihr firmenspezifisches Humankapital bedroht.[655] Zum anderen beugt Selbstüberschätzung persönlichen Zukunftszweifeln vor, die eine Versuchung (*moral hazard*) zu opportunistischem Endspielverhalten begründen könnten.[656] All das könnte dazu beitragen, dass die gravierenden Kosten, die die Selbstüberschätzung von Unternehmensleitern sporadisch verursacht, durch kontinuierliche Vorteile im Unternehmensalltag womöglich ausgeglichen werden.[657]

---

[650] *Beckhaus* in: Birk u.a., Unternehmenskauf 2010, 11, 37.

[651] *Langevoort* in: Gigerenzer/Engel, Heuristics and Law 2006, 87, 94; früher schon *Langevoort*, U Penn L Rev 1997, 101, 155: “Firms with ‘can-do’ cultures will thereby generate higher levels of internal effort and, by projecting self-confidence, be more successful in attracting external resources.”; vgl. aber ebd., 145: “we can see an optimistic culture as two-headed. On the one hand, it may be very useful to the firm as a motivator. On the other hand, optimism has a dark side […]”.

[652] *Paredes*, Fla St U L Rev 2005, 673, 698 ff. spricht vom “capitalized value of the reputational assets a management team builds when confidence and commitment are credibly projected” (701); vgl. auch *Cohen/Bailey*, J Mgmt 1997, 239, 276: “Perceptions can have stronger effects than objective reality in determining performance.”

[653] *Gully/Incalcaterra u.a.*, J Appl Psy 2002, 819 (Metastudie über 67 Primärstudien); *Kozlowski/Ilgen*, Psy Sci Publ Interest 2006, 77, 90 f.; *Mathieu/Maynard u.a.*, J Mgmt 2008, 410, 426; vgl. auch *Sniezek*, Org Behav Hum Dec Proc 1992, 124, 150: „groups are often preferred over the individual in situations with great uncertainty and a higher need for quality. One possible reason is […] enhanced confidence.“

[654] *Langevoort*, U Penn L Rev 1997, 101, 155 Fn. 192: “an optimistic culture or subculture may be an agency-cost reduction mechanism”.

[655] Vgl. schon oben bei Fn. 587; *Paredes*, Fla St U L Rev 2005, 673, 682, 739; *Langevoort* in: Gigerenzer/Engel, Heuristics and Law 2006, 87, 94; *Homberg/Osterloh*, JfB 2010, 269, 290 bei Fn. 21 m. Verw. auf einen Modellierungsversuch.

[656] *Paredes*, Fla St U L Rev 2005, 673, 694: “when a manager believes he is in a final period (that is, that he will be ousted from his job unless he has a major success), he may reasonably conclude that he has little to lose by taking a big risk”; ebenso *Langevoort*, U Penn L Rev 1997, 101, 155; vgl. auch *Langevoort*, Geo L J 2001, 797, 809.

Die vorliegende Arbeit muss es bei dieser Problemdarstellung bewenden lassen. Weder die Kosten noch der Nutzen, die ein Unternehmen von der Selbstüberschätzung seiner Führungsriege haben könnte, sind ohne Weiteres quantifizierbar. Zudem ist fraglich, inwieweit die ökonomische Analyse des Rechts überhaupt *normative* Orientierung bieten kann. Die hier aufgeworfene Frage, ob das Recht Urteilsverzerrungen – insbesondere jene der Selbstüberschätzung – korrigieren *sollte*, muss daher einer ausführlicheren Betrachtung vorbehalten bleiben. Gegenstand der vorliegenden Ausführungen war primär die Frage, ob das Recht durch das Kollegialprinzip solche Urteilsverzerrungen korrigieren *kann*. Daran bestehen erhebliche empirische Zweifel.

4. Gruppenentscheidung, Organisationspflichten und Haftungsmaßstab

Schließlich hat ein Verständnis der empirischen Befunde über kollegiales Verhalten auch Auswirkungen auf Haftungsfragen im Zusammenhang mit der Erfüllung gesellschaftsrechtlicher Organisations- und Verhaltenspflichten.[658] Etwa trifft den Vorstand nach § 91 II AktG die Pflicht, „geeignete Maßnahmen zu treffen […] damit den Fortbestand der Gesellschaft gefährdende Entwicklungen früh erkannt werden." Die Erfüllung dieser Pflicht setzt Wissen darüber voraus, welche Umstände empirisch dazu beitragen, dass Gefährdungen „erkannt werden". Dabei sind die oben (bei Fn. 608 und 611) schon angedeuteten menschlichen Neigungen zur Selbstrechtfertigung und Verleugnung widriger Informationen zu berücksichtigen, die durch die empirisch belegte Gruppenpolarisierung eher verstärkt als abgemildert werden dürften. Das könnte dafür sprechen, dass § 91 II AktG klar zugewiesene Einzelverantwortlichkeiten voraussetzt, statt etwa kollegiale Kontrollstrukturen.

Auch mit Blick auf § 93 I 2 AktG stellt sich die Frage, wann ein Vorstandsmitglied „vernünftigerweise annehmen" darf, in einer Gruppenentscheidung „auf der Grundlage angemessener Information" zu handeln.[659] Im Hinblick auf die US-amerikanische Rechtsprechung zu den Sorgfaltspflichten der Unternehmensleiter (*duty of care*) wurde dazu festgestellt:

---

[657] *Langevoort*, U Penn L Rev 1997, 101, 156: "While there are serious costs associated with ignoring danger signs in a small subset of cases, these costs may be outweighed by the profitability produced by the benign influences of organizational self-deception in others."

[658] Wegweisend *Fraidin*, UC Davis L Rev 2004, 1, passim, der eine gruppenpsychologische Fundierung der zersplitterten *duty-of-care*-Rechtsprechung vorschlägt: "it would make the law more coherent by describing the principles underlying the holdings in particular cases" (15).

[659] *Fraidin*, UC Davis L Rev 2004, 1, 16: "what behaviors characterize a group of people making a decision carefully?"

„Das Fallrecht bietet einem Juristen keine Orientierung, der entscheiden möchte, ob Unternehmensleiter ihre Sorgfaltspflichten dadurch erfüllen können, dass sie eine besonders lange Beratungssitzung abhalten statt den Rat eines externen Unternehmensberaters einzuholen. Haben Ratschläge eines Unternehmensberaters und lange Sitzungen vergleichbare Auswirkungen auf die Gruppenentscheidung? Das Fallrecht verfügt über kein Prinzip, das diese Frage beantworten könnte.“[660]

Nichts anderes gilt in Deutschland: Vorstandsmitglieder, die ihren Sorgfaltspflichten genügen wollen, bedürfen ebenso einer empirisch fundierten Handlungsanleitung wie die Richterin, die etwaige Haftungsfälle zu entscheiden hat (vgl. schon oben § 3 nach Fn. 37). Die in Gruppen festgestellten Polarisierungstendenzen könnten etwa erfordern, dass Gegenrede aktiv nachgesucht und wohlwollend aufgenommen werden muss. Bisweilen wird gar vorgeschlagen, den Widerspruchsgeist innerhalb des Organs durch Ernennung eines *advocatus diaboli* (*chief naysayer*) zu institutionalisieren.[661] Auch wenn dieser Vorschlag einstweilen vielleicht „rechtlich [.] noch nicht ausgereift“ erscheint[662] und empirisch genauer betrachtet werden müsste als es hier möglich ist,[663] könnte er auf lange Sicht durchaus den Sorgfaltsmaßstab des „ordentlichen und gewissenhaften Geschäftsleiters“ erweitern.

Gleiches gilt für den Aufsichtsrat (§ 116 S. 1 AktG), der sich bei Ausübung seiner Organisationsautonomie (§ 107 AktG) bekanntlich über die Frage Rechenschaft ablegen muss,

„wie er seine Arbeit zweckmäßigerweise einrichtet, um seinen gesetzlichen Funktionen und seiner Allgemeinverantwortung am besten gerecht zu werden, inwieweit er deshalb seine Aufgaben im Plenum erledigen oder einem Ausschuß übertragen will“.[664]

Die Entscheidung des Aufsichtsrats über die Ausgestaltung seiner Binnenstruktur ist mithin durch eine „Pflicht […] zu sachgerechter Organisation seiner Tätigkeit“[665] gebunden, weshalb bereits vor über vierzig Jahren prägnant resümiert wurde:

„Für die Beantwortung der Frage nach den Grenzen der inneren Organisation des Aufsichtsrats müssen zur Auslegung der aktienrechtlichen Vorschriften die Erkenntnisse der Sozialpsychologie über das gruppendynamische Verhalten herangezogen werden.“[666]

---

[660] *Fraidin*, UC Davis L Rev 2004, 1, 16.

[661] Ausf. *Paredes*, Fla St U L Rev 2005, 673, 740 ff.; *Fleischer*, ZGR 2011, 155, 161 f.; weitere Ansätze für “Dissent Through Process” in *Dorff*, Cardozo L Rev 2007, 2025, 2074 ff.

[662] So *Fleischer*, ZGR 2011, 155, 181.

[663] Dazu etwa die Metastudie von *Schwenk*, Org Behav Hum Dec Proc 1990, 161 und die spätere Studie von *Nemeth/Brown/Rogers*, Eur J Soc Psy 2001, 707.

[664] BGHZ 83, 106, 115; zust. *Habersack* in: Goette/Habersack, MK-AktG 2014, § 107 Rn. 93.

[665] *Drygala* in: Lutter/Schmidt, AktG 2010, § 107 Rn. 33.

[666] *Dose*, ZGR 1973, 300, 315.

III. Die rechtspolitische Aussagekraft der vorliegenden Untersuchung

Wie oben bereits vorausgeschickt (A.IV. a.E.), verfolgt die vorliegende Untersuchung kein rechtspolitisches Programm. Konkrete rechtspolitische Empfehlungen auf der Grundlage der oben gesichteten Empirie wären bestenfalls leichtfertig,[667] schlimmstenfalls fehlgeleitet. Dennoch kann empirische Forschung natürlich ebenso gut Veränderungen des Rechtszustands anregen wie sie auf die teleologische Auslegung innerhalb der Rechtsdogmatik einwirkt. Wenigstens drei rechtspolitische Fragen könnten von der oben gesichteten Empirie maßgeblich profitieren.

1. Kollegialprinzip als private Gestaltungsoption

Betrachtungen kollegialer Organtätigkeit dürfen nicht vergessen, dass Kollegialorgane, selbst wenn sie *faktisch* vielleicht den Normalfall darstellen (vgl. bei Fn. 62), *rechtlich* die absolute Ausnahme sind. Rechtliche Organpluralität, so hat der Rechtsvergleich ergeben (oben A.I.4.), ist fast nur in großen oder börsennotierten Publikumsgesellschaften sowie in Aufsichtsorganen anzutreffen. Stattdessen lautet die wohl häufigste Rechtsvorschrift über die Besetzung von Organen sinngemäß: „Das Organ muss aus einer oder mehreren Personen bestehen." Mithin ist private Gestaltungsfreiheit, also autonome Gestaltung des Rechts, der wichtigste rechtspolitische Anknüpfungspunkt für die empirische Forschung über das Kollegialprinzip.

Das gilt sowohl für Rechtsordnungen, die traditionell der kollegialen Entscheidung vertrauen, sie aber dispositiv stellen (vgl. § 76 II 2 AktG) als auch für Rechtsordnungen, die kollegialen Entscheidungen traditionell misstrauen, sie aber rechtlich ermöglichen[668] – ganz zu schweigen von agnostischen Rechtsordnungen, die die Führungsorganisation ganz der privatautonomen Regelung überlassen. Zudem ist selbst dort, wo die plurale Organbesetzung zwingend angeordnet ist, nicht selten die Einrichtung von Ausschüssen möglich und nötig. Soweit die jeweilige Rechtsordnung die Ausschussgröße in das Belieben des Organs stellt und beispielsweise die Delegation an einzelne Organvertreter ermöglicht (vgl. oben bei Fn. 169), ist die Gestaltungspraxis dann ebenso gefordert, eine im Einzelfall angemessene Geschäftsordnung zu entwerfen, die den empirisch belegten Nachteilen kollegialer Entscheidung Rechnung trägt.

Gerade wegen dieser Einzelfallausrichtung der Gestaltungspraxis ist es auch kaum möglich, ihr pauschale Rezepte an die Hand zu geben.[669] Für die kollegiale Besetzung eines Organs oder eines Ausschusses sprechen

---

[667] *Paredes*, Fla St U L Rev 2005, 673, 682: "because we still have a great deal to learn about decisionmaking, any recommendations for reforming corporate governance based on our present understanding of managerial psychology are tentative."

[668] Etwa Südkorea, vgl. oben bei Fn. 181 sowie das Zitat unten bei Fn. 676.

jenseits der hier betrachteten Argumente durchaus pragmatische Erwägungen (oben bei Fn. 300 und 301). Etwa können Organe je nach den Umständen des Unternehmens, der Branche und der Rechtsordnung als Repräsentationsforen der jeweiligen Bezugsgruppen dienen.[670] Das wird besonders am deutschen Aufsichtsrat deutlich, dessen Mitgliederzahl frühere Gesetzgeber bewusst nur behutsam eingegrenzt hatten,[671] weil zwar „zur reinen Ueberwachung eine geringere Anzahl genügend und fast eher geeignet" ist, zugleich aber „in der Praxis der Aufsichtsrat nicht lediglich der Ueberwachung dient, sondern auch dem Zweck einer Verbindung und Fühlung, z.B. mit Banken und Großlieferern bzw. -kreditgebern."[672] Dafür eignet sich natürlich nur eine plurale Organstruktur, und die Gestaltungspraxis mag gut beraten sein, alle sonstigen Nachteile des Kollegialprinzips dafür in Kauf zu nehmen. Zumindest *erkennen* muss sie diese Nachteile aber, um sie in Kauf nehmen zu können. Dann ist die Gestaltungspraxis in der (komfortablen) Lage, flexible Regelungen entwickeln zu können, die solche Vor- und Nachteile gegeneinander abwägen. Auch ohne pauschale Rezepte, welche Gestaltungen in welchen Fällen wie zu wählen wären, mag ihr deshalb mit einer Darstellung des empirischen Befundes bereits gedient sein.

2. CEO oder Vorstand – zum Streit um die beste Führungsstruktur

Weniger Flexibilität herrscht in der auf abstrakt-generelle Geltung angelegten Gesetzgebung. Dementsprechend verengt sich die Diskussion hier auf „Alles-oder-Nichts"-Lösungen, etwa hinsichtlich der Frage, ob rechtliche Organpluralität wünschenswert ist oder nicht. Diese Frage ist in der spezifischen Formulierung, ob das US-amerikanische Modell der Unternehmensführung – mit einem starken CEO (*chief executive officer*) an der

[669] *Eisenberg*, J Am Stat Assoc 2000, 665, 667: "The most important contribution of empirical legal research will not be statistical analyses that help determine individual cases."

[670] International verweisen etwa *Kakabadse/Kakabadse/Yang* in: Kakabadse/Kakabadse, Global Boards 2009, 96 auf "the importance of network ties in forging business links in China. […] The strength of the network was seen as providing access to scarce resources such as land, labour, raw materials, import and other licenses, insider information and endorsements. Thus, network ties were reported as still predominating over concerns of board performance efficiency. […] larger boards offered far better network accessibility."; ähnl. für die Schweiz *Böckli*, Schweizer AktR 2009, 1550.

[671] Anders etwa *IK Berlin*, DB 2000, 1573, 1578: „Der AR hat – soweit zulässig – sechs oder neun Mitglieder. […] Die Ausschüsse sollen mindestens drei und höchstens fünf Mitglieder haben."

[672] *Koenige/Teichmann/Koehler*, HGB 1936, § 86 AktG Anm. 2; *Klausing*, Begr. AktG-E 1937, 73; schon *Weber*, Wirtschaft 1922, 675 schrieb über die „in heutigen privatwirtschaftlichen bürokratischen Gebilden (Aktiengesellschaften) sich findenden kollegialen Kontrollinstanzen (Aufsichtsrat)": „normalerweise vereinigen diese Gebilde nicht Träger speziellen Fachwissens, sondern die ausschlaggebenden ökonomischen Hauptinteressenten".

Spitze – dem überkommenen deutschen Vorstandsmodell nicht möglicherweise überlegen ist, alles andere als neu. Schon der 34. Deutsche Juristentag 1926 beschäftigte sich ausführlich damit,[673] und zuletzt wurde das Thema mit der zunehmenden Globalisierung der Finanz- und Gütermärkte und einer in der Rechtspraxis oft wahrgenommenen Stärkung des Vorstandsvorsitzenden aktuell.[674] Die oben wiedergegebenen Befunde über Kollegialentscheidungen sind in diesem Zusammenhang natürlich sehr relevant, allerdings kann die Frage, welches Modell der Unternehmensführung überlegen ist, vor allem aus vier Gründen hier nicht beantwortet werden:

Erstens genügt kein noch so klarer *empirischer* Befund, um die *normative* Überlegenheit eines bestimmten Modells der Unternehmensführung zu begründen. Selbst wenn man aus der empirischen Sichtung schließen wollte, dass das Kollegialprinzip jede Wirkung verfehlt, die nach den fünf empirischen Grundannahmen intendiert war, wären umfangreiche Erörterungen dazu erforderlich, warum die Wirksamkeit der entscheidende *normative* Maßstab zur Beurteilung des Kollegialprinzips sein sollte. Womöglich ist die Unternehmensführung durch Kollegialorgane aus Gründen von Demokratie oder Gemeinwohl erforderlich, ohne dass es noch auf die fünf Annahmen zur Wirkungsweise ankäme.

Zweitens wurde vorliegend nur *ein* Modell der Unternehmensführung, nämlich das Kollegialprinzip, überhaupt angesprochen. Selbst wenn man die Wirksamkeit für normativ maßgeblich *und* das Kollegialprinzip für gänzlich unwirksam halten wollte, wäre damit noch nicht gesagt, dass das Direktorialprinzip in irgendeiner Hinsicht wirksamer wäre. Auch sind zahlreiche Führungsstrukturen denkbar, die zwischen den beiden Extremen liegen. Da jede augenscheinlich bessere Alternative auf empirisch haltlosen Annahmen beruhen könnte, müssten ihre Annahmen ebenso gründlich überprüft werden, bevor die Alternative rechtspolitisch in Betracht käme.

Drittens lässt sich aus der hier erfolgten Sichtung keineswegs schließen, dass das Kollegialprinzip seine Wirkung gänzlich verfehlt. Zum einen wurden nur fünf empirische Annahmen überprüft, die zwar zentral erscheinen, aber die vom Kollegialprinzip erhofften Wirkungen eben nicht abschließend wiedergeben. Zum anderen sind die empirischen Zweifel an einigen Grundannahmen bislang noch eher schwach und bedürfen weiterer empirischer Untersuchungen. Schließlich können selbst diejenigen Annahmen, die sich als empirisch sehr zweifelhaft herausgestellt haben, vielleicht so abgewandelt werden, dass sie mit dem empirischen Erkenntnisstand in Ein-

---

[673] Vgl. nur *von Hein*, ZHR 2002, 464, 473 f. m.w.N.

[674] Vgl. etwa *Bernhardt/Witt*, ZfB 1999, 825, 830 f. und die ausf. Beiträge von *Rieger* in: Lutter u.a., FS Peltzer 2001, 339, 47 ff.; *von Hein*, ZHR 2002, 464; *Hoffmann-Becking*, NZG 2003, 745; *Oesterle*, zfo 2003, 199; *Grundei*, BCCG DiskP 2004; *Wettich*, Vorstandsorganisation 2008, 295 ff.

klang kommen. Ein generelles Urteil über die Wirksamkeit des Kollegialprinzips erlaubt die vorliegende Untersuchung daher nicht.

Viertens schließlich – und hier kommen wir auf die äußere Gültigkeit zurück – beschränken sich die Unterschiede zwischen dem US-amerikanischen und dem deutschen Modell der Unternehmensführung nicht auf den einfachen Gegensatz zwischen Einzelentscheidung und Gruppenentscheidung. Vielmehr „gilt es, stets die gesamte *Governance-Konfiguration* im Auge zu halten“, also auch die im US-amerikanischen Recht stärker ausgeprägten Kontrollmechanismen, wie den weit entwickelten Kapitalmarkt und die jederzeitige Möglichkeit zur Abberufung des CEO.[675] Auch die unterschiedliche soziale Einbettung der Unternehmenstätigkeit und kulturelle Rahmenbedingungen (wie das Demokratieverständnis) wären bei der Frage nach der besten Führungsstruktur zu berücksichtigen, während sie vorliegend ausgeblendet blieben.

Wiewohl die hier zusammengestellten empirischen Erkenntnisse bei der rechtspolitischen Argumentation hilfreich sein mögen, erlauben sie also keine Antwort auf die Frage nach der „besten“ Führungsstruktur. Die festgestellten empirischen Zweifel an einzelnen Grundannahmen des Kollegialprinzips sprechen zwar gegen diese Annahmen, nicht aber gegen das Kollegialprinzip insgesamt oder gar *für* alternative Entscheidungsstrukturen. Dagegen zwingen sie mindestens dazu, das Kollegialprinzip auf eine verbesserte argumentative Grundlage zu stellen – und die zugrundeliegenden Annahmen transparent und dadurch überprüfbar zu machen.

Das betrifft freilich nicht nur den Vorstand, sondern auch die plurale Besetzung von Aufsichtsorganen, die im Rechtsvergleich zwar weit verbreitet, aber durchaus nicht alternativlos ist. Etwa lässt sich feststellen,

> „dass in Korea die Kollegialität eines Überwachungsorgans als ein die effiziente Überwachung verhinderndes Element angesehen wird. Dies wird […] damit begründet, dass die Aufsichtspersonen in einem Kollegialorgan weniger Verantwortungsgefühl haben“[676]

Gerade im Zuge der in den letzten Jahren vorangetriebenen Aufwertung und Professionalisierung des Aufsichtsrats wäre also eine fundierte Auseinandersetzung mit der Frage wünschenswert, worin eigentlich die plurale Besetzung des Organs ihre Rechtfertigung findet, und wie die Rolle des Aufsichtsratsvorsitzenden als exponiertem Einzelnen mit großem Einfluss

---

[675] *Grundei*, BCCG DiskP 2004, 1/15 (Zitat mit Hervorhebung im Original); *Hoffmann-Becking*, NZG 2003, 745, 746.

[676] *Yi*, Strukturreform 2004, 224 f., und weiter: „Zum anderen wird es damit begründet, dass […] eine Aufsichtsperson eine herrschende Stellung besitzen könnte und die anderen Aufsichtspersonen des Kollegialorgans allein dem Willen der beherrschenden Aufsichtsperson folgen könnten.“

innerhalb des Kollegialorgans im Lichte der hier dargestellten empirischen Erkenntnisse zu bewerten ist.[677]

3. Zum Wert demographischer Vielfalt (diversity)

Eine weitere rechtspolitische Frage, zu der die vorliegende Untersuchung beitragen kann, ist jene nach dem Wert demographischer Vielfalt in der Unternehmensführung (*diversity*). Im Deutschen Corporate Governance Kodex finden sich dazu die folgenden, erstmals Ende 2008 angekündigten,[678] Empfehlungen, deren Befolgung oder Nichtbefolgung nach § 161 AktG offenzulegen ist:

„Der Aufsichtsrat soll für seine Zusammensetzung konkrete Ziele benennen, die [… unter anderem] Vielfalt (Diversity) berücksichtigen." (Ziff. 5.4.1 S. 2 DCGK)

„Bei der Zusammensetzung des Vorstands soll der Aufsichtsrat auch auf Vielfalt (Diversity) achten und dabei insbesondere eine angemessene Berücksichtigung von Frauen anstreben." (Ziff. 5.1.2 S. 2 DCGK)

„Der Vorstand soll bei der Besetzung von Führungsfunktionen im Unternehmen auf Vielfalt (Diversity) achten und dabei insbesondere eine angemessene Berücksichtigung von Frauen anstreben." (Ziff. 4.1.5 DCGK).[679]

Diese „Einbeziehung aller Führungspositionen mittels eines Drei-Stufen-Plans" geht ganz allgemein von der „Notwendigkeit einer heterogenen Organstruktur" aus,[680] wobei nicht ganz klar ist, welche Art von Heterogenität sich der Kodexgeber vorstellt.[681] Auch die Europäische Kommission beschäftigte sich in ihren Grünbüchern zur „Corporate Governance in Finanzinstituten und Vergütungspolitik" und zum „Europäischen Corporate Governance-Rahmen" mit der Vielfalt in der Unternehmensführung.[682] Sie geht davon aus, der bisherige Mangel an demographischer Vielfalt habe erhebliche Kontrolldefizite verursacht[683] und hält Vielfalt daher für „eines

---

[677] Zum Aufsichtsratsvorsitzenden *Lutter/Krieger*, Rechte und Pflichten 2008, Rn. 656 ff.; allg. zur neuen Rolle des Aufsichtsrats noch *Lutter*, NJW 1995, 1133; *Lutter*, ZIP 2003, 737, 738; *Lutter*, DB 2009, 775, passim.

[678] *Ringleb/Kremer u.a.*, NZG 2010, 1161, 1163 (Rn. 661a).

[679] Kommentierung bei *Ringleb/Kremer u.a.*, NZG 2010, 1161; dazu nur *Sünner*, CCZ 2009, 185; *Mense/Rosenhäger*, GWR 2010, 311; *Weber-Rey/Handt*, NZG 2011, 1; *Ihrig/Meder*, ZIP 2012, 1210.

[680] *Weber-Rey/Handt*, NZG 2011, 1, 2 bzw. 5 mit der Erläuterung, dass ein „weites Meinungsspektrum die Gefahr einer zu einseitigen Sichtweise minimiert. Heterogen und vielfältig besetzte Kontrollgremien […] arbeiten kreativer und nachhaltiger als andere Gremien, die in festgefahrenen Denkmustern verharren" (2).

[681] *Sünner*, CCZ 2009, 185, 186; *Ihrig/Meder*, ZIP 2012, 1210, 1215.

[682] *Europäische Kommission*, Lessons 2010, 39 f., 43; *Europäische Kommission*, Grünbuch 2011, 6 ff.

der Schlüsselprobleme der Unternehmensverfassung“.[684] Organe mit größerer „Vielfalt in den Profilen der Mitglieder und ihres Werdegangs“, so die empirische Annahme, träfen „potentiell […] bessere Entscheidungen“.[685]

Zwar wurde die Förderung der Vielfalt als „Modethema“ kritisiert, das das Aktienrecht „mit gesellschaftspolitischen Vorstellungen überfrachtet und einer juristischen Analyse daher kaum zugänglich“ sei,[686] aus empirischer Sicht hingegen fallen solche rechtspolitischen Bemühungen auf sehr fruchtbaren Boden. Einerseits ist belegt, dass Vorstände traditionell (und mit starkem Vorsitzenden umso mehr) zu hoher demographischer Homogenität neigen;[687] das mag bei Aufsichtsräten oberflächlich nicht ganz so stark der Fall sein,[688] doch unterliegen auch sie den oben (bei Fn. 649) besprochenen Selektionsprozessen im Wirtschaftsleben, die bestimmte Charaktereigenschaften und Grundeinstellungen begünstigen. Andererseits hat die obige empirische Sichtung zahlreiche Zusammenhänge ergeben, in denen die Homogenität der Gruppe einen wichtigen Katalysator für negative gruppendynamische Prozesse bildete. So führte ein höherer Grad an geteilten Informationen zur Entstehung verborgener Informationsprofile und erschwerte zugleich die Erweiterung des Alternativenspektrums.[689] Ebenso setzt die Gruppenpolarisierung einen mindestens moderaten Gruppenkonsens voraus (dazu vor Fn. 545) und ist die anfängliche Verteilung von indi-

---

[683] *Europäische Kommission*, Lessons 2010, 9: “a lack of diversity of views within boards [.] may in some cases have contributed to the failure of non-executive board members to effectively challenge management decisions.”

[684] *Europäische Kommission*, Lessons 2010, 8: “one of key issues of corporate governance”.

[685] *Europäische Kommission*, Grünbuch 2011, 6; noch deutlicher *Europäische Kommission*, Lessons 2010, 9: “Diversity, not just of gender but also of race and social background, and the presence of employee representatives, broadens the debate within the boards and helps, as some say ‘to avoid the danger of narrow *group think*’, to which boards drawn from a narrow social range are prone.”

[686] *Fleischer*, ZGR 2011, 155, 157.

[687] *Westphal/Zajac*, Admin Sci Q 1995, 60; *Cox/Munsinger*, L Ctmp Prob 1985, 3/83, 105 f. m.w.N.: “continue to be distinguished by their homogeneity”; *Dorff*, Cardozo L Rev 2007, 2025; 2038 f.: “Public company boards overwhelmingly consist of white, middle-aged men from privileged backgrounds who have spent their careers working for large corporations.”; zur nationalen Vielfalt aber *Peltzer*, NZG 2011, 961, 962 Fn. 1.

[688] *Sünner*, CCZ 2009, 185, 187; *Forbes/Milliken*, Acad Mgmt Rev 1999, 489, 497 f.: “Contemporary boards […] represent a variety of industries and functional backgrounds, as well as significant numbers of lawyers, investment bankers, academics, and nonprofit executives who represent diverse educational and industry backgrounds.”

[689] M.a.W. *Stasson/Bradshaw*, Small Grp Res 1995, 296, 304: “nonoverlapping knowledge or unique expertise contributed to an aggregate level assembly effect”, daher bedürfe es “knowledge in complementary areas of expertise.” (306); *Drewes/Schultze/Schulz-Hardt* in: Frey/Bierhoff, Interaktion und Gruppe 2011, 221, 238: „Meinungsvielfalt ist also günstig“.

viduellen Einstellungen ein maßgebliches Kriterium für die kollektive Selbstüberschätzung (dazu vor Fn. 595).

In all diesen Fällen dürften die beschriebenen gruppendynamischen Effekte durch einen hohen Grad an kognitiver Heterogenität unter den Mitgliedern vermindert werden.[690] Das spricht für die im juristischen Schrifttum geäußerte Ansicht, die Gruppenmitglieder werden sich umso besser ergänzen, „je verschiedenartiger der Entwicklungsgang der Arbeitsteilnehmer ist. Denn damit wird das Feld umso schmaler, in dem sich die Fähigkeiten, Kenntnisse und Erfahrungen aller einzelnen decken."[691] Speziell im Gesellschaftsrecht ist die kognitive Heterogenität von Kollegialorganen daher wichtig.[692]

Damit ist freilich noch nicht gesagt, dass solche *kognitive* Heterogenität durch jedwede *demographische* Heterogenität erreichbar ist.[693] Zur geschlechtsbezogenen Heterogenität (*gender diversity*) beispielsweise liegen mittlerweile mindestens drei Metastudien und mehrere Forschungsberichte vor, die hier nicht gesichtet werden müssen, weil das jüngst an anderer Stelle geleistet wurde. Dort ergab sich, „dass die einschlägige Forschung derzeit keine Schlüsse auf eine generelle ökonomische Vorteilhaftigkeit von Gender Diversity zulässt" – was aber auch daran liegen kann, dass „Studien im Kontext deutschsprachiger Unternehmen ausgesprochen selten" sind.[694]

Jedenfalls lassen sich die empirischen Befunde wiederum nicht ohne Weiteres in normative Empfehlungen übersetzen. Soweit man mit der Kommission davon ausgeht, dass „gute" – sprich: gründlich reflektierte –

---

[690] *Homberg/Osterloh*, JfB 2010, 269, 290: „Verschiedene Studien zeigen, dass insbesondere kognitive Diversität (d.h. z.B. Diversität in Ausbildung und Funktion) die Entscheidungsqualität in schwierigen Situationen erhöht."; früher schon *Fraidin*, UC Davis L Rev 2004, 1, 53 m.w.N.: "Many laboratory studies, as well as more naturalistic studies, support the causal connection between dissent and good group decisions." und *Sniezek*, Org Behav Hum Dec Proc 1992, 124, 133, 135: "Disagreement within a group provides the opportunity for enlarging the domain of information processed, e.g., consideration of a wider range of alternatives or con as well as pro reasons for any given alternative." und noch früher verwies schon *Leu/Werner* in: Bender, Tatsachenforschung 1972, 127, 142 darauf, "dass heterogene Gruppen unter günstigen Umständen bessere Leistungen erbringen können als […] homogen zusammengesetzte Gruppen."

[691] *Körner*, ZöR 1937, 55, 57; „Daher beurteilt die Menge […] besser, der eine diese, der andere jene Seite an denselben und alle zusammen das Ganze", so *Aristoteles* zit. nach *Dagtoglou*, Kollegialorgane 1960, 21 f. mit Fn. 64.

[692] Zutr. *Grundei/von Werder*, AG 2005, 825, 829; *Dallas*, San Diego L Rev 2003, 781, 814 f.; vgl. auch *Forbes/Milliken*, Acad Mgmt Rev 1999, 489, 496 f. – wenn auch bezogen auf Gruppendenken statt Gruppenpolarisierung.

[693] Ausf. diff. *Larson*, Synergy 2010, 314 ff., auch zur Unterscheidung von äußerer ("surface") und innerer ("deep") *diversity.*

[694] *Boerner/Keding/Hüttermann*, zfbf 2012, 37.

Entscheidungen die normative Zielgröße bilden, bietet die vorliegende empirische Sichtung zwar starke Argumente für eine größere kognitive Vielfalt im Kollegialorgan. Zugleich könnte diese aber mit Nachteilen verbunden sein, die vorliegend nicht hinreichend reflektiert wurden. Die Kommission selbst erwähnt etwa den höheren Zeitaufwand, den die Koordinierung von Organmitgliedern mit sehr verschiedenen Hintergründen verursachen dürfte.[695] Weiterhin birgt eine Besetzung des Kollegialorgans mit sehr unterschiedlichen Persönlichkeiten ein nicht ganz unerhebliches internes Konfliktpotential.[696] Auch die maßgeblich auf Vertrauen beruhende „soziale Dynamik" eines Organs könnte durch zu große demographische Vielfalt gestört werden,[697] und die Ausbildung eines transaktiven Gedächtnisses, das die Informationsverarbeitung in der Gruppe erleichtern könnte (oben bei Fn. 459), wird umso schwieriger, je vielfältiger das Organ besetzt ist.

Womöglich handelt es sich bei all diesen Einwänden eher um quantitative als um qualitative („Vielfalt ist gut, aber mehr Vielfalt ist nicht immer besser.") aber um das zu beurteilen, wäre eine vertiefte empirische Auseinandersetzung erforderlich. Einstweilen lässt sich nur feststellen, dass rechtspolitische Bestrebungen wie jene der Europäischen Kommission aufgrund der hier vorgelegten empirischen Sichtung einige Plausibilität beanspruchen können.

E. Fazit: Das Kollegialprinzip hält nicht, was es (bisher) verspricht.

Die vorliegende Untersuchung hat fünf empirische Annahmen, die der Geltung des Kollegialprinzips im Gesellschaftsrecht zugrunde liegen, anhand des empirischen Erkenntnisstands untersucht. Keine der Annahmen ließ sich in ihrer bisherigen Form erhärten. Was das Kollegialprinzip bisher zu versprechen schien, kann es also empirisch nicht halten. Während rechts-

---

[695] *Europäische Kommission*, Grünbuch 2011, 6; ebenso *Moye*, Business 2004, 166: "more diverse points of view expressed always complicate the process of reaching a decision."

[696] *Peltzer*, NZG 2011, 961 f.; *Forbes/Milliken*, Acad Mgmt Rev 1999, 489, 497 f.: "diversity is a 'double-edged' sword for groups […] it is also associated with higher levels of conflict, interaction difficulties, and lower levels of integration." m.w.N.; *Hollingshead/Wittenbaum u.a.* in: Poole/Hollingshead, Small Groups 2005, 21, 41: "Diversity that increases task-relevant skills and knowledge can be helpful to performance. However, too much of that diversity may not be helpful because of group members' failure to understand one another. Diversity on characteristics not related to the task may be problematic in that it may elicit negative emotional reactions or attitudes and increase turnover."

[697] *Langevoort*, Geo L J 2001, 797, 831.

politische Implikationen dieses Befunds denkbar, aber nicht das Ziel der vorliegenden Untersuchung waren, empfiehlt sich zumindest eine Schärfung der dogmatischen Argumentation. Einige dogmatische Konsequenzen der empirischen Sichtung wurden aufgezeigt, langfristig sollten die zur Rechtfertigung des Kollegialprinzips herangezogenen Annahmen transparenter vermittelt und empirisch besser belegt werden.

§ 5

Zusammenfassung in Thesen

„Eine kritische Rechts- und Sozialphilosophie wird […] versuchen, die Erkenntnisse der Realwissenschaften zur kritischen Durchleuchtung des geltenden Rechts heranzuziehen“[1]
– Hans Albert 1971

„Wer sich nicht mehr hinter empirischen Unwägbarkeiten verstecken kann, muss sich zu Werten bekennen. […] Ergo, und ironischerweise, wächst mit der Menge empirischen Wissens auch die Zahl und Qualität rechtlicher Wertentscheidungen.“[2]
– Michael Saks 1989

1. *Evidenzbasierte Jurisprudenz* ist die Integration von juristischer Lebenserfahrung und empirischer Erkenntnis. Eine Evidenzbasierung der juristischen Forschung und Praxis ermöglicht – wie in der Medizin in den letzten zwanzig Jahren – sachgerechtere Entscheidungen.

a. Die Kluft zwischen Sollen und Sein ist ernstzunehmen, aber durch *konditionale Empfehlungen* überbrückbar. Schließlich kommt die hermeneutische Wertungsjurisprudenz trotz ihrer dogmatischen Grundausrichtung nicht ohne empirisches Tatsachenwissen aus.

b. Empirisches Tatsachenwissen können Juristen entweder in der Rolle des *Produzenten* selbst erheben – was aus praktischen, nicht aber grundsätzlichen, Erwägungen selten geschieht – oder in der Rolle des *Rezipienten* aus der empirischen Sozialforschung beziehen.

c. Damit die Rechtswissenschaft dieses Tatsachenwissen berücksichtigen kann, verfügt sie über drei wichtige *Schnittstellen*: Dogmatik (Wortlaut bzw. Normzweck), Rechtspolitik (Gesetzesvorbereitung bzw. Vollzugs- und Folgenkontrolle) sowie Metajurisprudenz.

---

[1] *Albert* in: Vernunft 2011, 163, 180.

[2] *Saks*, zit. nach *Schuck*, J L Edu 1989, 323, 335: “When one can no longer hide behind empirical uncertainties, one is compelled to state one’s normative preferences. […] Ergo, and somewhat ironically, the more positive knowledge there is, the more and better the law’s normative work will be.”; ähnl. *Nard*, Wake Forest L Rev 1995, 347, 361: “empirical data […] serves a democratic purpose. Confronted with a societal problem, our officials, assuming they desire to address the issue, are forced to divulge their political and legal philosophy.”

d. Im *Gesellschaftsrecht* ist empirisches Tatsachenwissen besonders gefragt und seine Vermehrung als Zukunftsaufgabe anerkannt. Die empirische Gesellschaftsrechtsforschung dominieren bislang in Deutschland die Rechtstatsachenforschung (als eine empirische Richtung unter vielen möglichen) und in den USA die Ereignisstudienforschung.

2.
Die Rezeption empirischer Forschung setzt ein gründliches Verständnis des empirischen *Forschungsprozesses* voraus.

a. Quantitative empirische Forschung folgt in allen Disziplinen demselben Grundschema, das sich in *sieben Phasen* unterteilen lässt. Großes Gewicht kommt vor allem der gründlichen Studienplanung zu, sowie der Veröffentlichungsphase als maßgeblichem Nadelöhr.
b. Evidenzbasierte Jurisprudenz steht oft einem unaufschiebbaren *Entscheidungsinteresse* gegenüber. Juristen sind deshalb auf ihre aus anekdotischer Evidenz geronnene Lebenserfahrung angewiesen, wenn und soweit es an relevanter empirischer Erkenntnis fehlt.
c. *Empirische Erkenntnis* ist die Summe der kumulativen Evidenz und anders als Lebenserfahrung durch ein intersubjektiv nachvollziehbares und replizierbares Erhebungsverfahren gewonnen. Empirische Erkenntnis ist für allgemeine Feststellungen zuverlässiger als Lebenserfahrung.
d. Quantitativ-empirische Forschung beruht grundsätzlich auf *statistischer Auswertung*. Statistik ist eine durch soziale Normen und Konventionen geprägte rhetorische Übung und für Juristen mit der nötigen Einarbeitung leicht anschlussfähig.
e. Die *Veröffentlichung* empirischer Evidenz ist strukturell gefiltert und zugunsten trügerischer Scheinevidenz verzerrt. Es bedarf daher systematischer Forschungsauswertungen (Metastudien), um verlässliche empirische Erkenntnis zu gewinnen.

3.
Eine kritisch reflektierte Untersuchung des empirischen Forschungsprozesses ergibt sieben *Rezeptionsregeln*, die eine verlässliche Evidenzbasierung der Jurisprudenz sicherstellen:

a. *Alle empirische Forschung ist implizit normativ.* An vielen Stellen des Forschungsprozesses finden sich Annahmen und Wertungen, die der Rezipient berücksichtigen muss.
b. *Sorgfältige Planung geht vor statistischer Raffinesse.* Die entscheidende Hürde für verlässliche empirische Erkenntnis ist die gründliche Planung, nicht die statistische Auswertung.
c. *Ohne Hypothese kein Ergebnis.* Zur Planung gehört die Benennung von Hypothesen, ohne die schließende Statistik wertlos ist. Im Zweifel bedarf jede Erkenntnis der Replikation.

d. *Ergebnis ist nicht die Signifikanz, sondern die Effektgröße.* Signifikanztests werden weltweit missverstanden und haben nichts mit der praktischen Relevanz eines Ergebnisses zu tun.
e. *Ein Bild sagt mehr als tausend Signifikanzen.* Grafische Darstellungen erlauben oft eine bessere und wissenschaftlich angemessenere Würdigung von Daten als deren statistische Auswertung.
f. *Statistik darf keine Magie, muss aber MAGIC sein.* Statistik (und Wissenschaft insgesamt) beruht auf rhetorischer Überzeugungskraft, die mindestens fünf einzelne Faktoren umfasst.
g. *Synthese vor Primärstudien, Forschungsauswertung vor Forschungsbericht.* Empirische Erkenntnis entsteht kumulativ, daher gehört empirische Forschung nur dort aus Primärstudien rezipiert, wo Forschungsberichte fehlen, und aus diesen nur dort, wo Forschungsauswertungen fehlen. Das ist die Erkenntnishierarchie der Evidenzbasierung.

4.
Ein überzeugendes Beurteilungskriterium für die Güte und den Nutzen empirischer Forschung bilden die *Gültigkeiten*, nach denen sich empirische Erkenntnisse unterscheiden lassen.
a. *Innere Gültigkeit* misst die Aussagekraft einer Studie zur eindeutigen Feststellung von Ursachenzusammenhängen. Solche Eindeutigkeit erfordert eine Kovariation von Ursache und Wirkung, die Nachzeitigkeit der Ursache und das Fehlen von Verunreinigungen.
b. *Äußere Gültigkeit* misst die Übertragbarkeit einer Erkenntnis auf den für die Forschungsfrage relevanten Lebenssachverhalt. Diese beruht entweder auf wertendem Vergleich nach der Übertragungslehre oder (seltener) auf der Verallgemeinerung aus Zufallsstichproben.
c. Praktisch lässt sich Eindeutigkeit nur durch gezielte *Experimente* sicherstellen, Übertragbarkeit nur durch die möglichst berührungsfreie *Beobachtung* eines Lebenssachverhaltes. Folglich stehen sie in einem Spannungsverhältnis zueinander.

5.
Das Gültigkeitenkriterium erlaubt es, innerhalb des allgemeinen Sieben-Phasen-Schemas quantitativ-empirischer Forschung einzelne *Erhebungsmethoden* zu unterscheiden.
a. Die höchste Übertragbarkeit und geringste Eindeutigkeit finden sich in *Korrelationsstudien*, das umgekehrte Extrem belegen *Laborexperimente*. Dazwischen liegt ein Spektrum, innerhalb dessen die beobachtenden Methoden nahtlos in die experimentellen übergehen.
b. Innerhalb des *Spektrums* finden sich vor allem Feldstudien, Befragungen, Vignettenstudien, Quasi-Experimente und Feldexperimente, die je eigene Abwägungen der Gültigkeiten treffen und deshalb für eine Forschungsfrage besser oder schlechter geeignet sein können.

c. Labor- und Feldmethoden erbringen mehrheitlich *qualitativ* konsistente Ergebnisse, können aber je nach den Umständen des Einzelfalls *quantitativ* voneinander abweichen.
d. Die Gesamtschau von sieben unterschiedlichen Studien sowie einer Metastudie zur Frage, ob höhere *Erfolgsvergütung* die Leistung von Vorstandsmitgliedern steigert, deutet vorläufig auf kontraproduktive Effekte der variablen Vorstandsvergütung hin.
e. Diese empirische Erkenntnis erhöht *rechtsdogmatisch* die Begründungslast für die Einführung solcher Vergütungssysteme nach § 87 I 1 AktG, zieht *rechtsökonomisch* die These in Zweifel, dass variable Vergütung Interessenkonflikte reduziert, und spricht *rechtspolitisch* gegen Ziff. 4.2.3 DCGK und ähnliche Bemühungen des Gesetzgebers.

6.
Die *Rezeption* empirischer Erkenntnis durch Juristen ist vor allem dort gewinnbringend, wo rechtliche Institutionen auf klaren empirischen Annahmen beruhen, zu denen umfangreiche empirische Forschung existiert. Das ist im Gesellschaftsrecht etwa beim Kollegialprinzip der Fall.
a. Das *Kollegialprinzip* ist kein Wertungsprinzip, sondern ein funktionsbestimmter Rechtsbegriff, der in Deutschland zum Bestand des allgemeinen Organisationsrechts gehört. Er bildet einen gemeinsamen Nenner von privaten Verbandsorganen, Richter- und Verwaltungsgremien.
b. Das Kollegialprinzip umfasst *zwei verschiedene* Grundsätze, die bisher vermengt wurden: Organpluralität (Organe werden mit mehreren Entscheidungsträgern besetzt.) und Organkollektivität (Mehrere Entscheidungsträger gehen durch grundsätzlich gemeinsame und gleichberechtigte Beschlusstätigkeit jeweils gleiche Verpflichtungen ein.)
c. *International* sind vielfältige Variationen feststellbar: Unter zwanzig verglichenen Rechtsordnungen findet sich keine, die das Kollegialprinzip nicht kennt, aber auch keine, in der Verbandsorgane ausnahmslos plural besetzt und kollektiv ausgestaltet sein müssen.
d. In den verschiedenen deutschen Rechtsgebieten und ausländischen Rechtsordnungen wird die Geltung des Kollegialprinzips maßgeblich auf fünf *empirische Grundannahmen* gestützt, zu deren Überprüfung mindestens sechs empirische Forschungsrichtungen beitragen können, darunter vor allem die sozialpsychologische Kleingruppenforschung.
e. Die Annahme, das Kollegialprinzip erhöhe die *Motivation* der Entscheidungsträger, ist für kognitiv anspruchsvolle oder komplexe Aufgaben sowie in Beschlusssituationen, wie sie in gesellschaftsrechtlichen Organen bestehen, sehr zweifelhaft.
f. Die Annahme, das Kollegialprinzip erziele *Synergien* durch Verbreiterung der Informationsgrundlage, ist vor allem in den Fällen zweifel-

haft, wo es darauf ankäme, weil die beste Handlungsoption verborgen ist – hier verengt die Gruppendiskussion den Blick eher.

g. Die Annahme, das Kollegialprinzip *entschleunige* die Informationssuche, ist auf Grundlage erster Studien eher zweifelhaft. Danach scheinen Gruppen nicht weniger vorschnell zu entscheiden als Einzelne, allerdings bedarf es hier wohl noch weiterer Forschung.
h. Die Annahme, das Kollegialprinzip führe zu gegenseitiger Überwachung und *Moderation*, ist sehr zweifelhaft. Im Gegensatz tendieren Gruppen eher dazu, die Meinungen ihrer Mitglieder noch zu polarisieren und Eskalationsneigungen zu verstärken.
i. Die Annahme, das Kollegialprinzip könne die *Selbstüberschätzung* von Unternehmensleitern abmildern, erscheint angesichts überzeugender Studien eher zweifelhaft. Auch diese Zweifel sind aber sehr vorläufig, da sie auf wenigen Studien beruhen.
j. Mangels empirischer Absicherung der genannten Annahmen lassen sich die bisherigen *rechtsdogmatischen* Zweckerwägungen nicht aufrecht erhalten. *Rechtspolitisch* mag der empirische Befund vor allem erneutes Nachdenken über die Eignung von CEO- und Vorstandsstrukturen sowie die Vorteile demographischer Vielfalt anregen.

7.

Evidenzbasierte Jurisprudenz ist die Vision einer sachgerechteren und methodisch reflektierteren Rechtswissenschaft. Sie wird die Entwicklung des Gesellschaftsrechts nachhaltig prägen.

Anhang: Empirie in Kirchners Bibliographie 1965–91

In der vom Bibliotheksleiter des Bundesgerichtshofs Hildebert Kirchner für die Jahre 1965–1991 in 38 Fortsetzungen veröffentlichten „Bibliographie zum Gesellschaftsrecht“[1] sind 34 Monographien und 49 Aufsätze nachgewiesen, deren Titel empirische Methoden erwähnen.[2]

Die tatsächliche Zahl der empirisch ausgerichteten Arbeiten dürfte jedoch deutlich höher sein. Zum Beispiel lassen nur drei der zwanzig in jener Bibliographie nachgewiesenen Arbeiten von Herbert Hansen – alle rechtstatsächlich – ihre empirische Ausrichtung ausdrücklich im Titel erkennen. In diesem Fall wurden die 17 übrigen Aufsätze von Hand nachgetragen, zahlreiche ähnlich gelagerte Fälle könnten allerdings übersehen worden sein.

In alphabetischer Sortierung verzeichnete Kirchner die folgenden 100 Veröffentlichungen:[3]

BALZ, GERHARD K.: Die Beendigung der Mitgliedschaft in der GmbH. Eine empirische u. dogmatische Untersuchung zur Ausschließung u. zum Austritt von Gesellschaftern. - Berlin: Duncker u. Humblot 1984. 232 S. (Schriften z. Wirtschaftsrecht. 49.)

BALZ, GERHARD K.: Rechtstatsachen zur Ausschließung und zum Austritt von Gesellschaftern der GmbH. In: GmbH-Rundschau. 1983, 8. S. 185-193

BAUCHOWITZ, HANS: Die Lageberichtspublizität der Deutschen Aktiengesellschaft. Eine empir. Unters, zum Stand d. Berichterstattung gem. § 160 Abs. l AktG. - Frankfurt a. M., Bern, Las Vegas: Lang 1979. III, 240 S. (Europ. Hochschulschriften. Reihe 5, Volks- u. Betriebswirtsch. Bd. 215.) Zugl. München, Betriebswirtschaft! Diss. 1978.

---

[1] 971 Seiten, von *Kirchner*, ZGR 1972, 104 bis *Kirchner*, ZGR 1992, 648.

[2] Volltextsuche nach den Wortstämmen **empir** (53), **tats(a/ä)ch** (23), **statist** (6), **umfrag** (5), **zahlen** (2), **fakten** (2), **quantit** (1), **erhebung** (1), **befrag** (0), **eviden** (0), **daten** (0), abzgl. zehn Doppeltreffern.

[3] Zitate unkorrigiert und ungeprüft übernommen aus den im Internetangebot des Verlags Walter de Gruyter verfügbaren und durch OCR-Texterfassung (*optical character recognition*) erschlossenen Folgen von Kirchners Bibliographie.

BAUMANN, HORST, WILHELM REISS: Satzungsergänzende Vereinbarungen – Nebenverträge im Gesellschaftsrecht. Eine rechtstatsächliche und rechtsdogmatische Untersuchung. In: ZGR. 1989, 2. S. 157-215.

BAUMANN, ROLAND: Abfindungsregelungen für ausscheidende Gesellschafter von Personenhandelsgesellschaften. Eine empir. Untersuchung. - Stuttgart, Rer. pol. Diss. 1987. XII, 338 S., S.XIII-XXV.

BOEMLE, MAX: Aktienrechtsrevision und Stille Reserven. Meinungen, Vorurteile, Tatsachen. In: Schweizer. Aktiengesellschaft. 1979, 3/4. S. 84—109

COENENBERG, ADOLF GERHARD: Jahresabschlußinformation und Kapitalmarkt. Zur Diskussion empirischer Forschungsansätze und -ergebnisse zum Informationsgehalt von Jahresabschlüssen für Aktionäre. In: Schmalenbachs Zs. f. betriebswirtschaftl. Forschung. 1974, N. F. H. 10. S. 647-657

DAMM, GÜNTHER, JOSEF HALBINGER und GEORG KELLINGHUSEN: Die Berichterstattung über Methodenänderungen gemäß § 160 Abs. 2 AktG. — Eine Erhebung anläßlich des Übergangs zur Teilwertberechnung der Pensionsverpflichtungen. In: Schmalenbachs Zs. f. betriebswirtschaftl. Forschung. 1977, 6. S. 367-375

DELLINGHAUSEN, RÖTGER FRHR. VON: Innere Strukturen der GmbH u. Co. KG. Rechtstatsachen und rechtliche Beurteilung. - Münster, Rechtswiss. Diss. 1980 (1981). XXXIV, 237 S

Die BEDEUTUNG von Einigungsstellen im Rahmen der Betriebsverfassung. Empirische Analyse der Wirkungsweise und Funktionsfähigkeit v. Einigungsstellen verfahren. T. 1.2. - Bamberg, Stuttgart: Poeschel.
1. Walter A. Oechsler und Michael Kerzel: Theoret. Grundlagen. 1984. 76S., 2. Anl.
2. Walter O. Oechsler und Thorleif Schönfeld: Ergebnisse empirischer Untersuchungen. 1985. 104 S. (Bamberger betriebswirtschaftliche Beiträge. 36. 44. DBW-Depot 86-4-4. 86-4-5.)

FEUERSTACK, RAINER: Unternehmenskonzentration. Theoret. Grundbegriffe u, empirische Ergebnisse für d. Bundesrep. Deutschland 1960-1970. - Neuwied: Luchterhand 1975. IV, 234 S. (Luchterhand economics.)

GERUM, ELMAR, BERND RICHTER und HORST STEINMANN: Unternehmensführung und Konzernmitbestimmung. Empirische Befunde zu den formalen Einflußstrukturen in mitbestimmten konzernverbundenen Aktiengesellschaften. In: Die Mitbestimmung. 1982, 7. S. 247-250.

GERUM, ELMAR, BERND RICHTER und HORST STEINMANN: Zur Steuerung öffentlicher Konzerne. Empirische Befunde zur institutionellen Durchsetzung des öffentl. Interesses in Aktiengesellschaftskonzernen. In: Zs. f. öffentl. u. gemeinwirtschaftl. Unternehmen. 1986, 3. S. 279-296.

GERUM, ELMAR, BERND RICHTER, und HORST STEINMANN: Unternehmenspolitik im mitbestimmten Konzern. Empirische Befunde zur Ausgestaltung v. Einflußstrukturen in mitbestimmten konzernverbundenen Aktiengesellschaften. In: Die Betriebswirtschaft. 1981, 3. S.345-360,

GERUM, ELMAR: Mitbestimmung und Effizienz. Neuere theoretische Entwicklungen u. empirische Befunde. In: Unternehmensverfassung. Festschrift f. Erich Potthoff. Baden-Baden: Nomos Verl. Ges. 1989. S. 46-60.

GMBH-TASCHENBUCH. Das Wichtigste zum Gesellschafts- und Steuerrecht der GmbH in ABC-Form, hrsg. von d. Centrale für GmbH Dr. Otto Schmidt. Lexikon: Dieter Schulze zur Wiesche. Formulare: Karl Eder. Statistik: Herbert Hansen. - Köln: O. Schmidt 1981. XIV, 306 S.

GORALCZYK, DIETMAR: Der Internationale Konzern. Zu Genesis, Funktionsweise und Empirie e. modernen Kapitalform. — Frankfurt a. M., Bern, Cirencester/U. K.: P. D. Lang 1979. 437 S. (Europ. Hochschulschriften. Reihe 5: Volks- u. Betriebswirtschaft. Bd. 229.)

GRASSLEY, MARY JANE: Gesetzliche Grundlagen und rechtstatsädiliche Erscheinungsformen in der Bekämpfung der Monopole und der Unternehmenskonzentration im Antitrustrecht der USA. Unter Berücksichtigung der Rechtsprechung d Obersten Bundesgerichtes. - Mainz, Rechts- u. wirtschaftswiss. Diss. 1966. VI, 113 S.

GROSS, MICHAEL: Das Innenverhältnis der GmbH. Eine rechtstatsächliche Untersuchung der höchstrichterlichen Rechtsprechung. - Bielefeld, Diss. 1987. XXII, 165 S

GRUHLER, WOLFRAM: Die Eigenkapitalausstattung der deutschen Unternehmen. Fakten, Auslandsvergleich, Ursachen und Folgerungen. In: Die Aktiengesellschaft. 1977, 3. S. 73-76

HAAS, HELMUT, und JOACHIM OSTHOFF: Konzernrechnungslegung in der Versicherungswirtschaft. Eine empirische Untersuchung von Geschäftsberichten der Jahre 1967 bis 1970. In: Die Wirtschaftsprüfung. 1973, 21. S. 573—582; 1973, 22. S. 597—604

HAEGER, BERND: Das System „Rechnungsführung und Statistik (RuSt)" in der DDR. In: Der Betrieb. 1990, 22, DDR-Report 1. S. 3015-3016.

HAHN, JÜRGEN: Jahresabschlußprüfung bald für 50 000 Unternehmen? GmbH-Umfrage bringt Zahlen. In: Der Betrieb. 1981, 49. S.2461-2462.

HALBINGER, JOSEF: Erfolgsausweispapier. Eine empirische Untersuchung z. Bilanzpolit. Verhalten deutscher Aktiengesellschaften. - Berlin: E. Schmidt 1980. XIX, 230 S. (Betriebswirtschaftl. Studien. 39.)

Hans-Böckler-Stiftung. - RICHTER, BERND: Der mitbestimmte Aktiengesellschaftskonzern. Eine ökonomisch-empirische Analyse zum Mitbestimmungsgesetz 1976. - Köln: Bund-Verl. 1983. IX, 394 S. l.

HANSEN, HERBERT: Das wirtschaftliche Gewicht der Gesellschaften m.b.H. In: GmbH-Rundschau. 1978, 6. S. 126-128

HANSEN, HERBERT: Die Aktiengesellschaften im Jahre 1970. Eine Auswertung statistischer Unterlagen. In: Die Aktiengesellschaft. 1971, 5. S. 148-149

HANSEN, HERBERT: Die Entwicklung der GmbH. In: GmbH-Rundschau. 1977, 7. S. 145-146

HANSEN, HERBERT: Die Entwicklung der Kapitalgesellschaft 1973. In: Die Aktiengesellschaft. 1974, 7. S. 218-220

HANSEN, HERBERT: Die Entwicklung der Kapitalgesellschaften 1971. Eine statistische Auswertung der Handelsregistereintragungen. In: Die Aktiengesellschaft. 1972, 9. S. 271-273

HANSEN, HERBERT: Die Entwicklung der Kapitalgesellschaften 1972. In: Die Aktiengesellschaft, 1973, 7. S. 219-221

HANSEN, HERBERT: Die Entwicklung der Kapitalgesellschaften und anderer Publizitätspflichtiger Unternehmen. In: Die Aktiengesellschaft. 1975, 10. S. 270- 276

HANSEN, HERBERT: Die Gesellschafter der größeren GmbH. In: GmbH-Rundschau. 1977, 10. S. 217-218

HANSEN, HERBERT: Die Gesellschafter größerer GmbH. In: GmbH-Rundschau. 1980, 5. S.99-102

HANSEN, HERBERT: Die GmbH gewann weiter an Bedeutung. In: GmbH-Rundschau. 1985, 7. S.209-212.

HANSEN, HERBERT: Die GmbH gewinnen weiter an Gewicht. In: GmbH-Rundschau. 1987,2. S. 50-52.

HANSEN, HERBERT: Die Kapitalausstattung der GmbH. In: GmbH-Rundschau. 1981, 11. S.259-262.

HANSEN, HERBERT: Die Rechnungslegung der Konzerne. In: Die Aktiengesellschaft. 1971, 9. S. 281—283

HANSEN, HERBERT: Entwicklung, Bilanzstruktur und Geschäftsergebnisse der Aktiengesellschaften. In: Die Aktiengesellschaft. 1973, 8. S. 253-258

HANSEN, HERBERT: Entwicklung, regionale Gliederung und Struktur der GmbH. In: GmbH-Rundsch. 1988, 1. S. 15-18.

HANSEN, HERBERT: Jahresabschlüsse 1976. In: Die Aktiengesellschaft. 1977, 10. S. 285—288

HANSEN, HERBERT: Jahresabschlußergebnisse nach Wirtschaftszweigen und Rechtsformen. In: Die Aktiengesellschaft. 1977, S. 189-192

HANSEN, HERBERT: Neuere Ergebnisse über die Entwicklung der GmbH. In: GmbHRundschau. 1986, 2. S. 37-38.

HANSEN, HERBERT: Wachsende Bedeutung der Gesellschaften mbH. In: GmbHRundschau. 1981, 5. S. 101-104.

HAUSCHILDT, JÜRGEN, und HANS-W. KORTMANN: „Sonstige finanzielle Verpflichungen" (§ 285 Nr. 3 HGB) als Gegenstand der Berichterstattung - eine empirische Analyse. In: Die Wirtschaftsprüfung. 1990, 15. S. 420^25.

HORST, KARL-HEINZ: Geschäftsführung, Vertretung und Beschlußfassung bei Personenhandelsgesellschaften. Eine empirische Untersuchung d. Vertragspraxis. - Frankfurt a. M., Bern: Lang 1981. XLIX, 600 [20] S. (Europäische Hochschulschriften. Reihe 2, Rechtswiss. Bd. 265.)

HUNGER, JOE R.: Die Jahresabschlußprüfung im Meinungsbild der Wirtschaftsprüfer, ihrer Mandanten und der Öffentlichkeit. Ergebnisse einer empirischen Untersuchung. In: Betriebswirtschaftl. Forschung u. Praxis. 1980, 1. S. 21-33

JÄGER, WERNER, und MICHAEL VOGELGESANG: Konzernrechnungslegung von Aktienbanken. Eine empirische Untersuchung von Geschäftsberichten der Jahre 1968, 1969 und 1970. In: Die Wirtschaftsprüfung. 1973, 15. S. 389—400

KLEINHERNE, HERIBERT JOHANNES: Erscheinungsformen und Gestaltungsmöglichkeiten personenbezogener Aktiengesellschaften unter bes. Berücks. der neueren rechts tatsächlichen Entwicklung. - Berlin, FU, Diss. 1990. 155 S.

KNAAK, ROLAND: Firma und Firmenschutz. Eine rechtstatsächliche Unters, z. Praxis d. Kennzeichenrechts. — Köln (usw.): Heymann 1986. XI, 288 S

KORNBLUM, UDO: Rechtstatsachen zum Unternehmens- und Gesellschaftsrecht. In: GmbH-Rundschau. 1981, 10. S. 227-235.

KORNBLUM, UDO: Weitere Rechtstatsachen zum Unternehmens- und Gesellschaftsrecht. In: GmbH-Rundschau. 1983. 2. S. 29-35; 1983, 3. S. 61-70

KREKELER, HANS-DIRK, und HAGEN LICHTENBERG: Satzungsänderungen nach § 58 Abs. 2 AktG. Eine Rechtstatsachenuntersuchung. In: Zs. f. d. ges. Handelsrecht u. Wirtschaftsrecht, Bd. 135. 1971, 4. S. 362-373

KREMER, ARNOLD: Die GmbH als Rechtsform freiberuflicher Partnerschaften. Eine Rechtstatsachenstudie zur Lösung des Konflikts zwischen den Vereinigungsbedürfnissen der freien Berufe und den Wesensmerkmalen der juristischen Person. - Berlin, FU, Rechtswiss. Diss. 1980. XXI, 201 S

LIMBACH, JUTTA: Theorie und Wirklichkeit der GmbH. Die empir. Normaltypen d. GmbH u. ihr Verhältnis zum Postulat von Herrschaft u. Haftung. — Berlin: Duncker u. Humblot 1966. 128 S. (Sdiriftenreihe d. Inst. f. Rechtssoziologie und Rechtstatsachenforschung an der Freien Universität Berlin. Bd. 2)

LOIDL, VOLKER: Die GmbH ohne erwerbswirtschaftliche Zielsetzung. Eine rechtstatsächl. Untersuchung. Köln: Otto Schmidt 1970. XXVIII, 125 S. (Rechtsfragen d. Handelsgesellschaften. 26.)

MEIER, GERT: Genossenschaften als per-se-Kartelle? Ein Rechtstatsachenbeitrag aus dem Bereich der „Einkaufsgenossenschaften. In: Zs. f. d. ges. Handelsrecht u. Wirtschaftsrecht. Bd. 142. 1978, 2. S. 124-146

MITBESTIMMUNG als Kampfaufgabe. Grundlagen, Möglichkeiten, Zielrichtungen. Eine theoretische, ideologiekritische u. empirische Untersuchung zur Mitbestimmungsfrage in der Bundesrepublik. Hrsg. v. Inst. f. marxistische Studien u. Forschungen (IMSF) Frankfurt a. M. Red.: Johannes Henrich v. Heiseler [u. a.] - Köln: Pahl-Rugenstein 1971. 348 S. (Beiträge d. IMSF. 2.)

MITBESTIMMUNG. Probleme und Perspektiven der empirischen Forschung. Hrsg. u. eingel. von Hans Diefenbacher u. Hans G. Nutzinger. - Frankfurt a. M., New York: Campus-Verl. 1981. 403 S.

MÖLLER, WOLF-PETER: Der Erfolg von Unternehmenszusammenschlüssen. Eine empirische Untersuchung. - München: Minerva-Publikation 1983. 403 S. (Minerva-Fachserie Wirtschafts- u. Sozialwissenschaften.) Zugl. Diss., Saarbrücken, 1982.

MORCK, WINFRIED: Die vertragliche Gestaltung der Beteiligung an Personen-Handelsgesellschaften. Eine empirische Untersuchung. - Frankfurt a. M., Bern, Cirences-

ter/U.K.: P.D.Lang 1980. XXXVIII, 543, 18S. (Bürgerl. Recht, Handels- u. Verkehrs-recht. Bd.6.)

MÜLLER, HELMUT: Mitbestimmung im Aufsichtsrat und Kontrolle der Unternehmens-politik. Eine empir. Unters, z. Handhabung repräsentativer Mitbestimmung in e. Mit-arbeiterunternehmen. — Frankfurt a.M., Bern (usw.): P. Lang 1986. VIII, 333 S. (Europ. Hochschulschriften. Reihe 5: Volks- u. Betriebswirtschaft. 693.)

NAUJOKS, WILFRIED, RAINER PAUSCH: Kooperationsverhalten in der Wirtschaft. Eine empirische Untersuchung unter bes. Berücks. d. grenzüberschreitenden Kooperation. - Göttingen: Schwartz 1977. VIII, 139 S. (Beiträge z. Mittelstandsforschung. H. 32.)

NEUE Rechtstatsachen zum Unternehmens- und Gesellschaftsrecht. Von Udo Kornblum, Werner Kleinle, Roland Baumann, Bernhard Steffan. In: GmbH-Rundschau. 1985, 1. S. 7-18; 1985,2. S. 42-51.

NIEDENHOFF, HORST: 20 Jahre Montan-Mitbestimmung im Spiegel empirischer Untersuchungen. - Köln: Dt. Industrieverlags-GmbH 1971. 51 gez. Bl. (Berichte d. Dt. Industrieinstituts zu Gewerkschaftsfragen. 1971, 9.)

NIEMEIER, WILHELM GÜNTHER: Rechtstatsachen und Rechtsfragen der Einziehung voon GmbH-Anteilen. - Heidelberg, Hamburg: v. Decker 1982. XX, 391 S. (Hambur-ge:er Beiträge z. Handels-, Schiffahrts- u. Wirtschaftsrecht. Bd. 6)

PAUL, GÜNTER: Mitbestimmung im Aufsichtsrat. Eine vergleichende empirische Analyse. In: Mitbestimmung u. Arbeiterselbstverwaltung. Frankfurt a.M.: Campus-Verl. 1982. S. 299-343.

PELLENS, BERNHARD, ULRICH LINNHOFF: Auswirkungen der neuen Rechnungslegungsvorschriften auf die Unternehmensbeurteilung. Eine empirische Untersuchung. In: Die Wirtschaftsprüfung. 1989, 5. S. 128-138.

PERLITZ, MANFRED: Die Prognosefähigkeit von Kennzahlen aus Jahresabschlüssen und Kapitalflußrechnungen für das Wachstum von Unternehmen. Eine empirische Untersuchung deutscher Aktiengesellschaften. In: Schmalenbachs 2s. f. betriebswirtschaftl. Forschung. 1972,1. S. 1-21

PLEYER, KLEMENS, und ULF BRÜHMANN: Fakten und Formen gesellschaftlicher Kooperation der Kreditgenossenschaften. In: Festschrift f. Ludwig Schnorr von Carolsfeld. Köln 1972. S. 427-444

RICHTER, BERND: Konzernführung und Mitbestimmung. Zur empirischen Überprüfung kontroverser Literaturmeimmgen. In: Der Betrieb. 1983, 39. S. 2072-2075.

RICHTER, BERND: Umgehung der Konzernvorschriften des Mitbestimmungsgesetzes 1976 durch Widerlegung der Abhängigkeits- und/oder Konzern Vermutung? Einige empirische Befunde. In: Die Aktiengesellschaft. 1982, 10. S.261-268.

ROTH, GÜNTER H., MICHAEL LORBACHER und WOLFGANG MERKEL: Die Mitbestimmung des Arbeitnehmers in der Praxis - Ergebnisse einer empirischen Untersuchung. In: 2s. f. Unternehmens- u. Gesellschaftsrecht (ZGR). 1974, 3. S. 317-342

RÜHL, RALF: Rechtstatsachen zur Sachgründung im GmbH-Recht. - München: Florentz 1988. XVIII, 172 S. (Rechtswissenschaftl. Forschung u. Entwicklung. 145.) Zugl. Heidelberg, Diss.

SCHIFFELS, EDMUND W.: Der Aufsichtsrat als Instrument der Unternehmenskooperation. Eine empir. Untersuchung z. Zusammensetzung d. Aufsichtsrates aller Aktiengesellschaften in 15 Wirtschaftszweigen d. Verarbeitenden Gewerbes i. d. Bundesrep. Deutschland. - Frankfurt a.M.: R. G. Fischer 1981. V, 327 S. (Beiträge z. Angewandten Wirtschafts- u. Sozialforschung. Wirtschaftswissenschaften. 1.)

SCHILDBACH, THOMAS, MARKUS BEERMANN, und MICHAEL FELDHOFF: Lagebericht und Publizitätspraxis der GmbH. Eine empirische Untersuchung. In: BB. 1990, 33. S. 2297-2301.

SCHMIDLE, LOTHAR: Das Publizitätsverhalten deutscher Aktiengesellschaften im Erläuterungsbericht gem. § 160 Abs. 2 Satz 2-5 AktG. Eine empir. Untersuchung. - Frankfurt a. M., Bern: Peter D. Lang 1981. X, 213 S. (Beiträge z. Rechnungs-, Finanz- und Revisionswesen. Bd. 2.)

SCHREYÖGG, GEORG: Managerkontrolle als Problem der Unternehmensverfassung. Empirische Befunde zur Trennung von Eigentum und Verfügungsgewalt und ihre Implikationen. In: Unternehmensverfassung. Köln: Heymann 1983. S. 15.3-166.

SCHULTE, KARL-WERNER: Betriebliche Altersversorgung und neues Bilanzrecht. Ergebnisse einer empirischen Untersuchung. In: BB 1989, 10. S. 659-663.

SCHULTE, KARL-WERNER: Das Publizitätsverhalten der Börsenneulinge. Ergebnisse einer empirischen Untersuchung von Geschäftsberichten. In: BB 1987, 12. S. 786-792.

SCHULTE, KARL-WERNER: Die Berichterstattung über Rückstellungen für Vorruhestandsverpflichtungen. Ergebnisse einer empirischen Untersuchung. In: BB 1986, 14. S. 911-915.

SCHULTE, KARL-WERNER: Steuerliche Rechnungszinsfußerhöhung u. Handelsrechtl. Bilanzierung von Pensionsrückstellungen. Ergebnisse einer empirischen Untersuchung. In:BB 1985, 11. S. 702-710.

SCHULTE, KARL-WERNER: Steuerliche Richttafeländerung und handelsrechtliche Bilanzierung von Pensionsrückstellungen. Ergebnisse einer empirischen Untersuchung. In: BB 1986,35/36.5.2371-2375.

SPIE, ULRICH, und WOLFGANG BAHLMANN: Arbeitsdirektoren in der Eisen- u. Stahlindustrie. Ergebnisse einer empirischen Untersuchung. 1. Zwei verschiedene Arbeitsdirektoren. 2. Das Sozial- u. Ausbildungsprofil d. Arbeitsdir. 3. Das Erfahrungsprofil d. Arbeitsdir. In: Das Mitbestimmungsgespräch. 1978, 11. S. 272-278; 1978, 12. S. 301-308; 1979, 1. S. 10-18; 1979, 2. S. 37-44

SPRISSLER, WOLFGANG: Die Rechnungslegung von Einzelkaufleuten und Personenhandelsgesellschaften nach dem Publizitätsgesetz. Eine empir. Unters. - München, Staatswirtschaftl. Diss. 1976. VI, 228 S., 2 Bl.

STEINMANN, HORST, und ELMAR GERUM: Unternehmenspolitik in der mitbestimmten Unternehmung. Empirische Befunde zum Einfluß des Aufsichtsrates von Aktiengesellschaften. In: Das Mitbestimmungsgespräch. 1980, 2. S. 40-48

STEINMANN, HORST, und ELMAR GERUM: Unternehmenspolitik in der Mitbestimmten Unternehmung. Empirische Befunde zum Einfluß des Aufsichtsrates von Aktiengesellschaften. In: Die Aktiengesellschaft. 1980, 1. S. 1-10

STRECK, MICHAEL: Die Praxis der Registergerichte angesichts der Publizitätsverweigerung kleiner GmbH. Erkenntnisse aus einer impulse-Umfrage. In: GmbH-Rdsch. 1991, S. 407^09.

TEULINGS, A. M. W.: Eine Verhandlungstheorie der Mitbestimmung. Ein empirischer Test anhand des niederländischen Systems betrieblicher Demokratie. In: Mitbestimmung u. Demokratisierung. Wiesbaden: Dt. Univ.-Verl. 1989. S. 91-121.

THEISEN, MANUEL R.: Die Aufgabenverteilung in der mitbestimmten GmbH. Eine empir. Unters, d. statutar. Regelungsstruktur nach d. Mitbestimmungsgesetz 1976. - Königstein/Ts.: Hanstein 1980. XXI, 264 S. (Schriften z. Wirtschafts- u. Arbeitsrecht. Bd. 10.) Zugl. Berlin, Freie Univ., Rechtswiss. Diss.

THEISEN, MANUEL R.: Gesetzestypische und tatsächliche Stellung der Unternehmensorgane in AG und GmbH. In: Unternehmensverfassung. Köln: Heymann 1983. S. 49-66.

VIEWEG, KLAUS: Normsetzung und -anwendung deutscher und internationaler Verbände. Eine rechtstatsächl. u. rechtl. Untersuchung unter bes. Berücks. d. Sportverbände. - Berlin: Duncker & Humblot 1990. 396 S. (Schriften zum Bürgerlichen Recht. 132.) Zugl. Münster, Habil.-Schr.

VOGEL, C. WOLFGANG: Aktienrecht und Aktienwirklichkeit. Organisation und Aufgabenteilung von Vorstand und Aufsichtsrat. Eine empirische Untersuchung deutscher Aktiengesellschaften. - Baden-Baden: Nomos Verlagsges. 1980. 307S. (Schriften d. Vereinigung f. Rechtssoziologie. Bd. 3.) Zugl. Diss., Gießen.

Voss, RÜDIGER v.: Zahlen zur Mitbestimmung. Die betroffenen Unternehmen und ihre Arbeitnehmer. In: Gesellschaftspolitische Kommentare. 1974, 4. S. 44-48

WINKELMANN, MICHAEL: Aktienbewertung in Deutschland. - Königstein/Ts.: Hain 1984. XII, 242 S. (Quantitative Methoden der Unternehmungsplanung. 19.)

WINTER, HEINZ: Übersicht über die Ergebnisse der Umfrage zur rechtstatsächlichen Struktur der GmbH. In: GmbH-Rundschau. 1969, 6. S. 119—123; 1969, 7. S. 145—149

WITTE, EBERHARD : Der Einfluß der Arbeitnehmer auf die Unternehmenspolitik. Eine empirische Untersuchung. In: Die Betriebswirtschaft. 1980, 4. S. 541-559

ZACH, ROGER: Privatautonome Mitbestimmungsordnung auf Unternehmensebene — tatsächliche und rechtliche Beurteilung. In: Schweizerische Aktiengesellschaft. 1984, 4. S. 160-174.

ZOHLNHÖFER, WERNER: Unternehmenskonzentration - eine ökonomisch-technologische Notwendigkeit? Meinungen u. empirische Befunde. In: Der Bürger im Staat. 1973, 4. S. 236-240

ZUTT, JÜRG: Stimmbindungen gegenüber Dritten - Ergebnisse einer Umfrage. In: ZHR 1991, S.213-215.

Literaturverzeichnis

ABA, Business Law Section: Model Nonprofit Corporation Act, 3. Aufl., New York 2009.

ABA, Business Law Section: Model Business Corporation Act Annotated, 4. Aufl., New York 2011.

Abdel-Khalik, Rashad: On the Efficiency of Subject Surrogation in Accounting Research, The Accounting Review 1974 = Bd. 49, S. 743–750.

Abelson, Robert: Statistics As Principled Argument, (Reprint 2009), Hillsdale 1995.

Achterberg, Norbert: Die Evidenz als Rechtsbegriff, Die Öffentliche Verwaltung 1963, S. 331–339.

Adair, John: The Hawthorne effect: A reconsideration of the methodological artifact, Journal of Applied Psychology 1984 = Bd. 69, S. 334–345.

Adams, Renée, Benjamin Hermalin & Michael Weisbach: The Role of Boards of Directors in Corporate Governance: A Conceptual Framework and Survey, Journal of Economic Literature 2010 = Bd. 48, S. 58–107.

Adomeit, Klaus & Guillermo Frühbeck: Einführung in das spanische Recht: das Verfassungs-, Zivil-, Wirtschafts- und Arbeitsrecht des Königreichs Spanien, 2. Aufl., München 2001.

Aguinis, Herman, Charles Pierce, Frank Bosco, Dan Dalton & Catherine Dalton: Debunking Myths and Urban Legends About Meta-Analysis, Organizational Research Methods 2011 = Bd. 14, S. 306–331.

Ahdieh, Robert: The Role of Groups in Norm Transformation: A Dramatic Sketch, in Three Parts, Chicago Journal of International Law 2005 = Bd. 6, S. 231–266.

Aiken, Lewis: Some Observations and Recommendations Concerning Research Methodology In the Behavioral Sciences, Educational and Psychological Measurement 1994 = Bd. 54, S. 848–860.

Aiman-Smith, Lynda, Steven Scullen & Steve Barr: Conducting Studies of Decision Making in Organizational Contexts: A Tutorial for Policy-Capturing and Other Regression-Based Techniques, Organizational Research Methods 2002 = Bd. 5, S. 388–414.

Albert, Hans: Erkenntnis und Recht. Die Jurisprudenz im Lichte des Kritizismus (1971), in: Kritische Vernunft und rationale Praxis, Tübingen 2011, S. 163–181.

Aldhous, Peter: Journal rejects studies contradicting precognition, NewScientist Life, Zugriff am 28.6.2012 unter http://www.newscientist.com/article/dn20447-journal-rejects-studies-contradicting-precognition.html.

Alexander, Cheryl & Henry Jay Becker: The Use of Vignettes in Survey Research, Public Opinion Quarterly 1978 = Bd. 42, S. 93–104.

Alexy, Robert: Theorie der juristischen Argumentation. Die Theorie des rationalen Diskurses als Theorie der juristischen Begründung, Frankfurt/M. 1983.

Allen, Mike & Raymond Preiss: Replication and Meta-Analysis: A Necessary Connection, Journal of Social Behavior and Personality 1993 = Bd. 8, S. 9–20.

Alm, James: A Call for Replication Studies, Public Finance Review 2010 = Bd. 38, S. 4–10.

Anderson, Craig, James Lindsay & Brad Bushman: Research in the Psychological Laboratory: Truth or Triviality?, Current Directions in Psychological Science 1999 = Bd. 8, S. 3–9.

Angrist, Joshua & Jörn-Steffen Pischke: Mostly Harmless Econometrics. An Empiricist's Companion, Princeton 2009.

Angrist, Joshua & Jörn-Steffen Pischke: The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics, The Journal of Economic Perspectives 2010 = Bd. 24, S. 3–30.

Anscombe, Francis: Graphs in Statistical Analysis, The American Statistician 1973 = Bd. 27, S. 17–21.

Appleton, David, Joyce French & Mark Vanderpump: Ignoring a Covariate: An Example of Simpson's Paradox, The American Statistician 1996 = Bd. 50, S. 340–341.

Ariely, Dan: Predictably Irrational. The Hidden Forces That Shape Our Decisions, New York 2008.

Ariely, Dan, Uri Gneezy, George Loewenstein & Nina Mazar: Large Stakes and Big Mistakes, The Review of Economic Studies 2009 = Bd. 76, S. 451–469.

Arkes, Hal, Caryn Christensen, Cheryl Lai & Catherine Blumer: Two methods of reducing overconfidence, Organizational Behavior and Human Decision Processes 1987 = Bd. 39, S. 133–144.

Arkes, Hal, Robyn Dawes & Caryn Christensen: Factors influencing the use of a decision rule in a probabilistic task, Organizational Behavior and Human Decision Processes 1986 = Bd. 37, S. 93–110.

Arlen, Jennifer & Eric Talley: Experimental Law and Economics, Cheltenham 2008.

Arlt, Marie-Agnes: Französische Aktiengesellschaft: monistisches und dualistisches System im Spannungsfeld der Corporate Governance, Diss., Wien 2006.

Arndt, Dominik: Empirie in den Rechtswissenschaften – Fluch oder Segen? Chancen und Risiken der Einbeziehung empirischer Ansätze in die rechtswissenschaftliche Forschung (Diplomarbeit), Universität Augsburg 2008, Zugriff am 27.6.2014 unter http://opus.bibliothek.uni-augsburg.de/opus4/frontdoor/index/index/docId/1190.

Arndt, Dominik: Sinn und Unsinn von Soft Law. Prolegomena zur Zukunft eines indeterminierten Paradigmas, Diss., Baden-Baden 2010.

Arnold, Arnd: Die Steuerung des Vorstandshandelns. Eine rechtsökonomische Untersuchung der Principal-Agent-Problematik in Publikumskapitalgesellschaften, Habil. 2006, München 2007.

Ars Legis e.V. (Hg.): Das Recht der Kapitalgesellschaften in Europa, Heidelberg 2007.

Ashton, Robert: Pressure and Performance in Accounting Decision Settings: Paradoxical Effects of Incentives, Feedback, and Justification, Journal of Accounting Research 1990 = Bd. 28, S. 148–180.

Ashton, Robert & Sandra Kramer: Students As Surrogates in Behavioral Accounting Research: Some Evidence, Journal of Accounting Research 1980 = Bd. 18, S. 1–15.

Atkinson, Donald, Michael Furlong & Bruce Wampold: Statistical significance, reviewer evaluations, and the scientific process: Is there a (statistically) significant relationship?, Journal of Counseling Psychology 1982 = Bd. 29, S. 189–194.

Atteslander, Peter: Methoden der empirischen Sozialforschung, 13. Aufl., Berlin 2010.

Atzmüller, Christiane & Peter Steiner: Experimental Vignette Studies in Survey Research, Methodology: European Journal of Research Methods for the Behavioral and Social Sciences 2010 = Bd. 6, S. 128–138.

Augsberg, Ino: Von einem neuerdings erhobenen empiristischen Ton in der Rechtswissenschaft, Der Staat 2012, S. 117–125.

Augsberg, Steffen: Kooperative Wissensgenerierung im Gesundheitsrecht. Zum Umgang der Sozialgerichte mit Evidenzbasierter Medizin, Gesundheitsrecht 2012, S. 595–601.

Auspurg, Katrin, Thomas Hinz & Stefan Liebig: Komplexität von Vignetten, Lerneffekte und Plausibilität im Faktoriellen Survey, Methoden — Daten — Analysen 2009 = Bd. 3, S. 59–96.

Auspurg, Katrin, Thomas Hinz, Stefan Liebig & Carsten Sauer: Auf das Design kommt es an. Experimentelle Befunde zu komplexen Settings in Faktoriellen Surveys, Sozialwissenschaftlicher Fachinformationsdienst „Methoden und Instrumente der Sozialwissenschaften" 2009, S. 23–39.

Austin, Robert, Harold Ford & Ian Ramsay: Company Directors: Principles of Law and Corporate Governance, Chatswood 2005.

Baer, Susanne: Interdisziplinäre Rechtsforschung. Was uns bewegt., in: Grundmann/ Kloepfer/Paulus/Schröder/Werle, Festschrift 200 Jahre Juristische Fakultät der Humboldt-Universität zu Berlin, Berlin 2010, S. 917–936.

Bainbridge, Stephen: Why a Board? Group Decisionmaking in Corporate Governance, Vanderbilt Law Review 2002 = Bd. 55, S. 1–55.

Baldwin, John & Gwynn Davis: Empirical Research in Law, in: Cane/Tushnet, The Oxford Handbook of Legal Studies, Oxford 2003, S. 880–900.

Bandiera, Oriana, Iwan Barankay & Imran Rasul: Incentives for Managers and Inequality among Workers: Evidence from a Firm-Level Experiment, The Quarterly Journal of Economics 2007 = Bd. 122, S. 729–773.

Bargh, John: Nothing in Their Heads, Psychology Today (The Natural Unconscious), Zugriff am 28.6.2012 unter http://www.psychologytoday.com/blog/the-natural-unconscious/201203/nothing-in-their-heads.

Barker, Roger: Explorations in Ecological Psychology, American Psychologist 1965 = Bd. 20, S. 1–14.

Barnard, Jayne: Institutional Investors and the New Corporate Governance, North Carolina Law Review 1991 = Bd. 69, S. 1135–1187.

Barnard, Jayne: Narcissism, Over-optimism, Fear, Anger, and Depression: The Interior Lives of Corporate Leaders, University of Cincinnati Law Review 2008 = Bd. 77, S. 405–430.

Barnlund, Dean: A comparative study of individual, majority, and group judgment, The Journal of Abnormal and Social Psychology 1959 = Bd. 58, S. 55–60.

Baron, Robert: So Right It's Wrong: Groupthink and the Ubiquitous Nature of Polarized Group Decision Making, Advances in Experimental Social Psychology 2005 = Bd. 37, S. 219–253.

Baron, Robert & Norbert Kerr: Group Process, Group Decision, Group Action, 2. Aufl., Buckingham 2003.

Bartsch, Daniel & Christoph Börner: Werteffekte strategischer Desinvestitionen – Eine empirische Untersuchung am deutschen Kapitalmarkt, Zeitschrift für betriebswirtschaftliche Forschung 2007 = Bd. 59, S. 2–34.

Bauen, Marc & Silvio Venturi: Der Verwaltungsrat. Organisation, Kompetenzen, Verantwortlichkeit, Corporate Governance, Zürich 2007.

Bauer, Max, Bernhard Eccher, Bernhard König, Josef Kreuzer & Heinz Zanon: Italienisches Zivilgesetzbuch (Codice Civile), Autonome Provinz Bozen, Südtirol: Abteilung Anwaltschaft des Landes, Zugriff am 9.10.2012 unter http://www.provinz.bz.it/anwaltschaft/download/ProvBZ_ZGB_Fassung_Stand_24_11_2010_de.pdf.

Bauman, Yoram & Elaina Rose: Selection or indoctrination: Why do economics students donate less than the rest?, Journal of Economic Behavior & Organization 2011 = Bd. 79, S. 318–327.

Baumann, Horst & Wilhelm Reiss: Satzungsergänzende Vereinbarungen – Nebenverträge im Gesellschaftsrecht. Eine rechtstatsächliche und rechtsdogmatische Untersuchung, Zeitschrift für Unternehmens- und Gesellschaftsrecht 1989 = Bd. 18, S. 157–215.

Baumeister, Roy: Choking under pressure: Self-consciousness and paradoxical effects of incentives on skillful performance, Journal of Personality and Social Psychology 1984 = Bd. 46, S. 610–620.

Baums, Theodor: Rechtstatsachen zu „Institutionelle Investoren und Publikumsgesellschaft“, in: Feddersen/Hommelhoff/Schneider, Corporate Governance. Optimierung der Unternehmensführung und der Unternehmenskontrolle im deutschen und amerikanischen Aktienrecht, Köln 1996, S. 324–330.

Baums, Theodor, Florian Drinhausen & Astrid Keinath: Anfechtungsklagen und Freigabeverfahren. Eine empirische Studie, Zeitschrift für Wirtschaftsrecht 2011 = Bd. 32, S. 2329–2352.

Baums, Theodor & Christian Fraune: Institutionelle Anleger und Publikumsgesellschaft – Eine empirische Untersuchung, Die Aktiengesellschaft 1995, S. 97–112.

Baums, Theodor, Astrid Keinath & Daniel Gajek: Fortschritte bei Klagen gegen Hauptversammlungsbeschlüsse? Eine empirische Studie, Zeitschrift für Wirtschaftsrecht 2007, S. 1629–1650.

Baums, Theodor & Markus König: Investmentfonds im Universalbankkonzern: Rechtstatsachen und aktuelle Reformfragen, in: Forster/Grunewald/Lutter/Semler, Festschrift für Bruno Kropff. Aktien- und Bilanzrecht, Düsseldorf 1997, S. 3–36.

Baums, Theodor & Erik Theissen: Banken, bankeigene Kapitalanlagegesellschaften und Aktienemissionen, Zeitschrift für Bankrecht und Bankwirtschaft 1999, S. 125–134.

Baums, Theodor, Hans-Gert Vogel & Maja Tacheva: Rechtstatsachen zur Beschlusskontrolle im Aktienrecht, Zeitschrift für Wirtschaftsrecht 2000, S. 1649–1655.

Bayer, Christian & Carsten Burhop: A Corporate Governance Reform as a Natural Experiment for Incentive Contracts, Schmalenbach Business Review 2008 = Bd. 60, S. 378–399.

Bayer, Christian & Carsten Burhop: Corporate governance and incentive contracts: Historical evidence from a legal reform, Explorations in Economic History 2009 = Bd. 46, S. 464–481.

Bayer, Walter: Die Aktiengesellschaft im Spiegel der Rechtstatsachenforschung, Jena 2007.

Bayer, Walter: Aktienrecht in Zahlen: Studien des Instituts für Rechtstatsachenforschung der Friedrich-Schiller-Universität Jena zum Recht der Aktiengesellschaft, AG Sonderheft, Köln 2010.

Bayer, Walter & Thomas Hoffmann: Aktuelle Rechtstatsachen zur Verbreitung der Aktiengesellschaft, Die Aktiengesellschaft 2009, S. R30–R32.

Bayer, Walter, Thomas Hoffmann & Tobias Sawada: Beschlussmängelklagen, Freigabeverfahren und Berufskläger. Ergebnisse einer im Auftrag des BMJ erstellten empirischen Studie und Überlegungen de lege ferenda, Zeitschrift für Wirtschaftsrecht 2012 = Bd. 33, S. 897–911.

Bayer, Walter, Thomas Hoffmann & Dirk Weinmann: Ereignisstudienmethodik (Methodische Ergänzungen zu Bayer/Hoffmann/Weinmann, ZGR 2007, 457, 479 bei Fn. 72), Universität Jena, Lehrstuhl Prof. Walter Bayer, Zugriff am 18.11.2012 unter www.rewi.uni-jena.de/rewimedia/Downloads/LS_Prof_Bayer/Ereignisstudienmethodik.pdf.

Bayer, Walter, Thomas Hoffmann & Dirk Weinmann: Kapitalmarktreaktionen bei Ankündigung des Rückerwerbs eigener Aktien über die Börse. Eine empirische Studie vor dem Hintergrund der reformierten Kapitalrichtlinie, Zeitschrift für Unternehmens- und Gesellschaftsrecht 2007 = Bd. 36, S. 457–479.

Baysinger, Barry & Henry Butler: Corporate Governance and the Board of Directors: Performance Effects of Changes in Board Composition, Journal of Law, Economics & Organization 1985 = Bd. 1, S. 101–124.

Bazerman, Max, Toni Giuliano & Alan Appelman: Escalation of commitment in individual and group decision making, Organizational Behavior and Human Performance 1984 = Bd. 33, S. 141–152.

Bazerman, Max & Don Moore: Judgment in managerial decision making, 7. Aufl., Hoboken 2009.

Bazerman, Max & Margaret Neale: Nonrational escalation of commitment in negotiation, European Management Journal 1992 = Bd. 10, S. 163–168.

Bebchuk, Lucian, Alma Cohen & Allen Ferrell: What Matters in Corporate Governance?, Review of Financial Studies 2009 = Bd. 22, S. 783–827.

Bebchuk, Lucian & Michael Weisbach: The State of Corporate Governance Research, Review of Financial Studies 2010 = Bd. 23, S. 939–961.

Bechtold, Stefan: Die Grenzen zwingenden Vertragsrechts. Ein rechtsökonomischer Beitrag zu einer Rechtsetzungslehre des Privatrechts, Tübingen 2010.

Beck-Bornholdt, Hans-Peter & Hans-Hermann Dubben: Is the Pope an alien?, Nature 1996 = Bd. 381, S. 730.

Beck, Michael & Karl-Dieter Opp: Der faktorielle Survey und die Messung von Normen, Kölner Zeitschrift für Soziologie und Sozialpsychologie 2001 = Bd. 53, S. 283–306.

Beckert, Ute: Personalisierte Leitung von Aktiengesellschaften, unter besonderer Berücksichtigung der Europäischen Aktiengesellschaft (SE), Diss. 2008, Baden-Baden 2009.

Beckhaus, Gerrit: Irrationale Entscheidungsprozesse bei Unternehmenstransaktionen – Erkenntnisse der Behavioral Economics als Grundlage für Strategien zur Vermeidung irrationalen Verhaltens, in: Birk/Bruse/Saenger, Forum Unternehmenskauf 2009: Aus dem Münsteraner Studiengang „Mergers & Acquisitions“, Baden-Baden 2010, S. 11–42.

Bédard, Jean: Expertise in auditing: Myth or reality?, Accounting, Organizations and Society 1989 = Bd. 14, S. 113–131.

Begg, Colin & Jesse Berlin: Publication Bias: A Problem in Interpreting Medical Data, Journal of the Royal Statistical Society. Series A (Statistics in Society) 1988 = Bd. 151, S. 419–463.

Behrendt, Jenny: Corporate Governance in Deutschland und Japan: die Rolle von Aktionärsklagen im Rechtsvergleich, Saarbrücken 2007.

Beilock, Sian & Thomas Carr: On the fragility of skilled performance: What governs choking under pressure?, Journal of Experimental Psychology: General 2001 = Bd. 130, S. 701–725.

Beilock, Sian & Thomas Carr: When High-Powered People Fail, Psychological Science 2005 = Bd. 16, S. 101–105.

Beller, Sieghard: Empirisch forschen lernen. Konzepte, Methoden, Fallbeispiele, Tipps, 2. Aufl., Bern 2008.

Ben-David, Itzhak, John R. Graham & Campbell R. Harvey: Managerial Overconfidence and Corporate Policies, National Bureau of Economic Research (NBER) Working Paper No. 13711, Zugriff am 14.3.2012 unter http://www.nber.org/papers/w13711.

Bénabou, Roland: Groupthink: Collective Delusions in Organizations and Markets, National Bureau of Economic Research (NBER) Working Paper No. 14764, Zugriff am 15.4.2012 unter http://www.nber.org/papers/w14764.

Benda, Ernst & Karl Krenzer: Demoskopie und Recht, JuristenZeitung 1972, S. 497–501.

Benicke, Christoph: Pflichten des Vermögensverwalters beim Investitionsprozess, Zeitschrift für Unternehmens- und Gesellschaftsrecht 2004, S. 760–812.

Berger, James & Donald Berry: Statistical Analysis and the Illusion of Objectivity, American Scientist 1988 = Bd. 76, S. 159–165.

Berkemann, Jörg: Die richterliche Entscheidung in psychologischer Sicht, JuristenZeitung 1971, S. 537–540.

Berkemann, Jörg: Rechtsfortbildung – Aspekte tatsächlichen Richterverhaltens, Kritische Vierteljahresschrift für Gesetzgebung und Rechtswissenschaft 1988, S. 29–56.

Bernhardt, Wolfgang & Peter Witt: Unternehmensleitung im Spannungsfeld zwischen Ressortverteilung und Gesamtverantwortung, Zeitschrift für Betriebswirtschaft 1999 = Bd. 69, S. 825–845.

Bertrand, Marianne, Esther Duflo & Sendhil Mullainathan: How Much Should We Trust Differences-In-Differences Estimates?, The Quarterly Journal of Economics 2004 = Bd. 119, S. 249–275.

Best, Lisa, Laurence Smith & Alan Stubbs: Graph use in psychology and other sciences, Behavioural Processes 2001 = Bd. 54, S. 155–165.

Bettenhausen, Kenneth: Five Years of Groups Research: What We Have Learned and What Needs to Be Addressed, Journal of Management 1991 = Bd. 17, S. 345–381.

Beutel, Frederick Keating: Die experimentelle Rechtswissenschaft. Möglichkeiten eines neuen Zweiges der Sozialwissenschaft, Berlin 1971.

Beyer, Andreas & Viktor Freystedt: Downgrading: Malus für Wechsler?, Die Aktiengesellschaft 2011, S. R80–R82.

Bhagat, Sanjai & Roberta Romano: Event Studies and the Law: Part I: Technique and Corporate Litigation, American Law and Economics Review 2002 = Bd. 4, S. 141–168.

Bhagat, Sanjai & Roberta Romano: Event Studies and the Law: Part II: Empirical Studies of Corporate Law, American Law and Economics Review 2002 = Bd. 4, S. 380–423.

Bickel, Peter, Eugene Hammel & William O'Connell: Sex Bias in Graduate Admissions: Data from Berkeley, Science 1975 = Bd. 187, S. 398–404.

Bilger, Stefan: Evidenz-basierte Medizin und Leitlinien, in: Comberg/Klimm, Allgemeinmedizin, 4. Aufl., Stuttgart 2004, S. 74–78.

Bilsky, Wolfgang, Anna Döring, Michael Janik, Denise Weßel-Therhorn & Klaus-Peter Dahle: Kriminalprävention, Forensische Psychiatrie, Psychologie, Kriminologie 2012 = Bd. 6, S. 69–70.

Bilz, Kenworthey: Defending the (Mis)use of Statistics in Law, Journal of Institutional and Theoretical Economics 2010 = Bd. 166, S. 194–198.

Binder, Jens-Hinrich: Regulierungsinstrumente und Regulierungsstrategien im Kapitalgesellschaftsrecht, Tübingen 2011.

Binder, John: The Event Study Methodology Since 1969, Review of Quantitative Finance and Accounting 1998 = Bd. 11, S. 111–137.

Black, Bernard: Bidder Overpayment in Takeovers, Stanford Law Review 1989 = Bd. 41, S. 597–660.

Black, Nick: Evidence based policy: proceed with care, British Medical Journal 2001 = Bd. 323, S. 275–278.

Blair, Margaret & Lynn Stout: A Team Production Theory of Corporate Law, Virginia Law Review 1999 = Bd. 85, S. 247–328.

Bleicher, Knut: Geschäftsführung und Aufsicht im internationalen Vergleich – Einsichten und Empfehlungen, Zeitschrift für betriebswirtschaftliche Forschung 1988 = Bd. 40, S. 930–941.

Bleicher, Knut, Diethard Leberl & Herbert Paul: Unternehmungsverfassung und Spitzenorganisation. Führung und Überwachung von Aktiengesellschaften im internationalen Vergleich, Wiesbaden 1989.

Blinder, Alan: Making Monetary Policy by Committee, International Finance 2009 = Bd. 12, S. 171–194.

Blinder, Alan & John Morgan: Are Two Heads Better than One? Monetary Policy by Committee, Journal of Money, Credit and Banking 2005 = Bd. 37, S. 789–811.

Blinder, Alan & John Morgan: Leadership in Groups: A Monetary Policy Experiment, International Journal of Central Banking 2008 = Bd. 4, S. 117–150.

Blume, Jeffrey & Jeffrey Peipert: What Your Statistician Never Told You about P-Values, The Journal of the American Association of Gynecologic Laparoscopists 2003 = Bd. 10, S. 439–444.

BMJ: Gesetze / Verordnungen alphabetisch sortiert, Bundesministerium der Justiz, Zugriff am 20.11.2012 unter http://www.gesetze-im-internet.de/aktuell.html.

Böckli, Peter: Schweizer Aktienrecht, 4. Aufl., Zürich 2009.

Boerner, Sabine, Hannah Keding & Hendrik Hüttermann: Gender Diversity und Organisationserfolg – Eine kritische Bestandsaufnahme, Zeitschrift für betriebswirtschaftliche Forschung 2012 = Bd. 64, S. 37–70.

Bogle, John: Reflections on CEO Compensation, The Academy of Management Perspectives 2008 = Bd. 22, S. 21–25.

Bolton, Gary, Axel Ockenfels & Ulrich Thonemann: Managers and Students as Newsvendors, Management Science 2012 = Bd. 58, S. 2225–2233.

Bond, Charles & Linda Titus: Social facilitation: A meta-analysis of 241 studies, Psychological Bulletin 1983 = Bd. 94, S. 265–292.

Bones, Arina: We Knew the Future All Along: Scientific Hypothesizing is Much More Accurate Than Other Forms of Precognition—A Satire in One Part, Perspectives on Psychological Science 2012 = Bd. 7, S. 307–309.

Bonner, Sarah, Reid Hastie, Geoffrey Sprinkle & Mark Young: A Review of the Effects of Financial Incentives on Performance in Laboratory Tasks: Implications for Management Accounting, Journal of Management Accounting Research 2000 = Bd. 12, S. 19–64.

Börsch-Supan, Axel & Jens Köke: An Applied Econometricians' View of Empirical Corporate Governance Studies, German Economic Review 2002 = Bd. 3, S. 295–326.

Bortz, Jürgen & Nicola Döring: Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler, 4. Aufl., Heidelberg 2006.

Bostrom, Ann: Risk Perceptions: "Experts" vs. "Lay People", Duke Environmental Law & Policy Forum 1997 = Bd. 8, S. 101–113.

Bradney, Anthony: The Place of Empirical Legal Research in the Law School Curriculum, in: Cane/Kritzer, The Oxford Handbook of Empirical Legal Research, Oxford 2010, S. 1025–1043.

Brandts, Jordi & Gary Charness: The strategy versus the direct-response method: a first survey of experimental comparisons, Experimental Economics 2011 = Bd. 14, S. 375–398.

Brauer, Markus: Die aktienrechtliche Beurteilung von „appreciation awards" zu Gunsten des Vorstands, Neue Zeitschrift für Gesellschaftsrecht 2004, S. 502–509.

Braun, Reiner, Horst Eidenmüller, Andreas Engert & Lars Hornuf: Unternehmensgründungen unter dem Einfluss des Wettbewerbs der Gesellschaftsrechte, Zeitschrift für das gesamte Handels- und Wirtschaftsrecht 2013 = Bd. 177, S. 131–148.

Brewer, James: Statistical Rules-of-Thumb, Florida Journal of Educational Research 1988 = Bd. 30, S. 5–14.

Brickley, James & Christopher James: The Takeover Market, Corporate Board Composition, and Ownership Structure: The Case of Banking, Journal of Law and Economics 1987 = Bd. 30, S. 161–180.

Bridges, David, Paul Smeyers & Richard Smith: Evidence-Based Education Policy. What Evidence? What Basis? Whose Policy?, Chichester 2009.

Brocato, Joe, Akhil Kumar & Kenneth Smith: Individual versus group spot price forecasting in the international petroleum market: A case study, Managerial and Decision Economics 1989 = Bd. 10, S. 13–24.

Brockhuis, Jörn & Christian Schnell: Gesellschaftsrecht in Polen, 2. Aufl., München 2002.

Brockner, Joel & Jeffrey Rubin: Entrapment in Escalating Conflicts. A Social Psychological Analysis, New York 1985.

Brodbeck, Felix, Rudolf Kerschreiter, Andreas Mojzisch & Stefan Schulz-Hardt: Group Decision Making Under Conditions of Distributed Knowledge: The Information Asymmetries Model, The Academy of Management Review 2007 = Bd. 32, S. 459–479.

Brodmann, Erich: Gewerbe- und Industrie-Kommentar Aktienrecht, Berlin 1928.

Brown, Rayna & Neal Sarma: CEO overconfidence, CEO dominance and corporate acquisitions, Journal of Economics and Business 2007 = Bd. 59, S. 358–379.

Bryden, David: Scholarship About Scholarship, University of Colorado Law Review 1992 = Bd. 63, S. 641–650.

Bu, Yuanshi: Einführung in das Recht Chinas, München 2009.

Bundesgerichtshof: Urt. v. 24.2.1954 – Az. II ZR 63/53, Entscheidungen des Bundesgerichtshofs in Zivilsachen 1954 = Bd. 12, S. 327–337.

Bundesgerichtshof: Urt. v. 12.12.1953 – Az. VI ZR 242/52, Lindenmaier-Möhring – Kommentierte BGH-Rechtsprechung 1954, S. 584–585 (§ 157 BGB Nr. 1).

Bundesgerichtshof: Urt. v. 9.2.1957 – Az. 2 StR 508/56, Entscheidungssammlung des Bundesgerichtshofes in Strafsachen 1957 = Bd. 10, S. 208–217.

Bundesgerichtshof: Urt. v. 23.10.1975 – Az. II ZR 90/73, Entscheidungssammlung des Bundesgerichtshofes in Zivilsachen 1975 = Bd. 65, S. 190–196.

Bundesgerichtshof: Urt. v. 25.2.1982 – Az. II ZR 123/81, Entscheidungssammlung des Bundesgerichtshofes in Zivilsachen 1982 = Bd. 83, S. 106–122.

Bundesgerichtshof: Urt. v. 15.10.1996 – Az. VI ZR 319/95, Entscheidungssammlung des Bundesgerichtshofes in Zivilsachen 1996 = Bd. 133, S. 370–383.

Bundesgerichtshof: Urt. v. 21.12.2005 – Az. 3 StR 470/04, Entscheidungssammlung des Bundesgerichtshofes in Strafsachen 2005 = Bd. 50, S. 331–346.

Bundesgerichtshof: Beschl. v. 19.7.2010 – Az. II ZB 18/09, Entscheidungssammlung des Bundesgerichtshofes in Zivilsachen 2010 = Bd. 186, S. 229–242.

Bundesverfassungsgericht: Urt. v. 11.11.1986 – Az. 1 BvR 713/83, 1 BvR 921/84, 1 BvR 1190/84, 1 BvR 333/85, 1 BvR 248/85, 1 BvR 306/85, 1 BvR 497/85, Entscheidungen des Bundesverfassungsgerichts 1986 = Bd. 73, S. 206–261.

Bundesverfassungsgericht: Urt. v. 28.5.1993 – Az. 2 BvF 2/90, 2 BvF 4/92, 2 BvF 5/92, Entscheidungen des Bundesverfassungsgerichts 1993 = Bd. 88, S. 203–366.

Bundesverfassungsgericht: Beschl. v. 11.10.1994 – Az. 1 BvR 337/92, Entscheidungen des Bundesverfassungsgerichts 1994 = Bd. 91, S. 148–176.

Bundesverfassungsgericht: Beschl. v. 6.12.2005 – Az. 1 BvR 347/98, Entscheidungen des Bundesverfassungsgerichts 2006 = Bd. 115, S. 25–51.

Bungert, Hartwin: Gesellschaftsrecht in den USA: eine Einführung mit vergleichenden Tabellen, 3. Aufl., München 2003.

Bürkle, Jürgen: Richterliche Alltagstheorien im Bereich des Zivilrechts. Mit einer Analyse amtsrichterlicher Urteile in Zivilsachen zu richterlichen Theorien für die Glaubhaftigkeitsbeurteilung von Zeugenaussagen, Diss. 1983, Tübingen 1984.

Bürkle, Jürgen: Der Stichentscheid im zweiköpfigen AG-Vorstand, Die Aktiengesellschaft 2012, S. 232–237.

Burman, Leonard, Robert Reed & James Alm: A Call for Replication Studies, Public Finance Review 2011 = Bd. 39, S. 190.

Burns, Kylie & Terry Hutchinson: The impact of "empirical facts" on legal scholarship and legal research training, The Law Teacher 2009 = Bd. 43, S. 153–178.

Burnstein, Eugene & Amiram Vinokur: Persuasive argumentation and social comparison as determinants of attitude polarization, Journal of Experimental Social Psychology 1977 = Bd. 13, S. 315–332.

Burnstein, Eugene, Amiram Vinokur & Yaacov Trope: Interpersonal comparison versus persuasive argumentation: A more direct test of alternative explanations for group-induced shifts in individual choice, Journal of Experimental Social Psychology 1973 = Bd. 9, S. 236–245.

Busenitz, Lowell & Jay Barney: Differences between entrepreneurs and managers in large organizations: Biases and heuristics in strategic decision-making, Journal of Business Venturing 1997 = Bd. 12, S. 9–30.

Bushman, Brad & Gary Wells: Narrative Impressions of Literature: The Availability Bias and the Corrective Properties of Meta-Analytic Approaches, Personality and Social Psychology Bulletin 2001 = Bd. 27, S. 1123–1130.

Bydlinski, Franz: System und Prinzipien des Privatrechts, Wien 1996.

Cadman, Brian, Sandy Klasa & Steve Matsunaga: Determinants of CEO Pay: A Comparison of ExecuComp and Non-ExecuComp Firms, The Accounting Review 2010 = Bd. 85, S. 1511–1543.

Cadsby, Bram & Elizabeth Maynes: Laboratory Experiments in Corporate and Investment Finance: A Survey, Managerial and Decision Economics 1998 = Bd. 19, S. 277–298.

Cafri, Guy, Jeffrey Kromrey & Michael Brannick: A Meta-Meta-Analysis: Empirical Review of Statistical Power, Type I Error Rates, Effect Sizes, and Model Selection of Meta-Analyses Published in Psychology, Multivariate Behavioral Research 2010 = Bd. 45, S. 239–270.

Camerer, Colin: Behavioral economics: Reunifying psychology and economics, Proceedings of the National Academy of Sciences 1999 = Bd. 96, S. 10575–10577.

Camerer, Colin & Robin Hogarth: The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework, Journal of Risk and Uncertainty 1999 = Bd. 19, S. 7–42.

Cane, Peter & Herbert Kritzer: The Oxford Handbook of Empirical Legal Research, Oxford 2010.

Carpenter, Siri: Psychology's Bold Initiative, Science 2012 = Bd. 335, S. 1558–1561.

Carr-Hill, Roy: Welcome? To the brave new world of evidence based medicine, Social Science & Medicine 1995 = Bd. 41, S. 1467–1468.

Carroll, Lewis: Sylvie and Bruno Concluded, London 1893.

Carsten, Gebhard & Friedrich Korkisch: Das schwedische Aktienrecht, 2. Aufl., Frankfurt/M. 1986.

CCH Asia: China Company Law Guide Bd. 2 (Gesetzestexte), The Hague 2005.

CH: The Companies (Model Articles) Regulations 2008, Companies House Statutory Instrument 2008 No. 3229, Zugriff am 27.6.2014 unter http://www.legislation.gov.uk/uksi/2008/3229/pdfs/uksi_20083229_en.pdf.

Chambliss, Elizabeth: When do Facts Persuade? Some Thoughts on the Market for "Empirical Legal Studies", Law and Contemporary Problems 2008 = Bd. 71, S. 17–39.

Charness, Gary & Matthias Sutter: Groups Make Better Self-Interested Decisions, Journal of Economic Perspectives 2012 = Bd. 26, S. 157–176.

Chatterjee, Arijit & Donald Hambrick: It's All about Me: Narcissistic Chief Executive Officers and Their Effects on Company Strategy and Performance, Administrative Science Quarterly 2007 = Bd. 52, S. 351–386.

Chen, Yan & Sherry Xin Li: Group Identity and Social Preferences, American Economic Review 2009 = Bd. 99, S. 431–457.

Cheung, Stephen & Stefan Palan: Two heads are less bubbly than one: team decision-making in an experimental asset market, Experimental Economics 2012 = Bd. 15, S. 373–397.

Chiotellis, Aristide & Wolfgang Fikentscher: Rechtstatsachenforschung – ein heute noch erfüllbares Programm?, in: dies., Rechtstatsachenforschung. Methodische Probleme und Beispiele aus dem Schuld- und Wirtschaftsrecht, Köln 1985, S. 1–7.

Chui, Wing Hong: Quantitative Legal Research, in: McConville/Chui, Research Methods for Law, Edinburgh 2007, S. 46–68.

Cleveland, William: Graphs in Scientific Publications, The American Statistician 1984 = Bd. 38, S. 261–269.

Coase, Ronald: The Nature of the Firm, Economica 1937 = Bd. 4, S. 386–405.

Cohen, Susan & Diane Bailey: What Makes Teams Work: Group Effectiveness Research from the Shop Floor to the Executive Suite, Journal of Management 1997 = Bd. 23, S. 239–290.

Cole, Stephen, Jonathan Cole & Gary Simon: Chance and consensus in peer review, Science 1981 = Bd. 214, S. 881–886.

Collins, Barry & Harold Guetzkow: A Social Psychology of Group Processes for Decision-Making, New York 1964.

Conradt, Larissa & Christian List: Group decisions in humans and animals: a survey, Philosophical Transactions of the Royal Society B: Biological Sciences 2009 = Bd. 364, S. 719–742.

Cook, R. J., B. M. Dickens & J. N. Erdman: Emergency contraception, abortion and evidence-based law, International Journal of Gynecology & Obstetrics 2006 = Bd. 93, S. 191–197.

Cooper, Harris, Larry Hedges & Jeffrey Valentine: The Handbook of Research Synthesis and Meta-analysis, New York 2009.

Cooper, Harris & Alison Koenka: The overview of reviews: Unique challenges and opportunities when research syntheses are the principal elements of new integrative scholarship, American Psychologist 2012 = Bd. 67, S. 446–462.

Cooper, Harris & Robert Rosenthal: Statistical versus traditional procedures for summarizing research findings, Psychological Bulletin 1980 = Bd. 87, S. 442–449.

Cooter, Robert: Maturing into Normal Science: The Effect of Empirical Legal Studies on Law and Economics, University of Illinois Law Review 2011, S. 1475–1483.

Coursol, Allan & Edwin Wagner: Effect of positive findings on submission and acceptance rates: A note on meta-analysis bias, Professional Psychology: Research and Practice 1986 = Bd. 17, S. 136–137.

Couzin, Iain, Jens Krause, Nigel Franks & Simon Levin: Effective leadership and decision-making in animal groups on the move, Nature 2005 = Bd. 433, S. 513–516.

Cox, James & Stephen Hayne: Barking up the right tree: Are small groups rational agents?, Experimental Economics 2006 = Bd. 9, S. 209–222.

Cox, James & Harry Munsinger: Bias in the Boardroom: Psychological Foundations and Legal Implications of Corporate Cohesion, Law and Contemporary Problems 1985 = Bd. 48, S. 83–135.

Creifelds, Carl & Klaus Weber: Rechtswörterbuch, 19. Aufl., München 2007.

Cremades, Javier: Gesellschaftsrecht in Spanien: eine Einführung mit vergleichenden Tabellen, 3. Aufl., München 2001.

Crocker, Jennifer & Lynne Cooper: Addressing Scientific Fraud, Science 2011 = Bd. 334, S. 1182.

Croson, Rachel: Why and How to Experiment: Methodologies From Experimental Economics, University of Illinois Law Review 2002, S. 921–945.

Cumming, Geoff: Inference by Eye: Pictures of Confidence Intervals and Thinking About Levels of Confidence, Teaching Statistics 2007 = Bd. 29, S. 89–93.

Cumming, Geoff & Sue Finch: Inference by Eye: Confidence Intervals and How to Read Pictures of Data, American Psychologist 2005 = Bd. 60, S. 170–180.

Cuñat, Vicente & Maria Guadalupe: Executive compensation and competition in the banking and financial sectors, Journal of Banking & Finance 2009 = Bd. 33, S. 495–504.

Curschmann, Jan & Claudio Jolowicz: Das neue brasilianische GmbH-Recht, GmbH-Rundschau 2003, S. 1185–1197.

Cycyota, Cynthia & David Harrison: What (Not) to Expect When Surveying Executives, Organizational Research Methods 2006 = Bd. 9, S. 133–160.

Dagtoglou, Prodromos: Kollegialorgane und Kollegialakte der Verwaltung, Diss. 1958, Stuttgart 1960.

Dallas, Lynne: The Multiple Roles of Corporate Boards of Directors, San Diego Law Review 2003 = Bd. 40, S. 781–820.

Dalton, Dan, Catherine Daily, Alan Ellstrand & Jonathan Johnson: Meta-analytic reviews of board composition, leadership structure, and financial performance, Strategic Management Journal 1998 = Bd. 19, S. 269–290.

Dauner-Lieb, Barbara: Unternehmerische Tätigkeit zwischen Kontrolle und Kreativität, in: Crezelius/Hirte/Vieweg, Festschrift für Volker Röhricht zum 65. Geburtstag – Gesellschaftsrecht, Rechnungslegung, Sportrecht, Köln 2005, S. 83–103.

Davies, Huw, Sandra Nutley & Peter Smith: What Works? Evidence-based policy and practice in public services, Bristol 2000.

Davies, Philip: What is Evidence-based Education?, British Journal of Educational Studies 1999 = Bd. 47, S. 108–121.

Davis, James: Group decision and social interaction: A theory of social decision schemes, Psychological Review 1973 = Bd. 80, S. 97–125.

Davis, James: Some compelling intuitions about group consensus decisions, theoretical and empirical research, and interpersonal aggregation phenomena: Selected examples 1950–1990, Organizational Behavior and Human Decision Processes 1992 = Bd. 52, S. 3–38.

Davis, Kevin: Legal Universalism: Persistent Objections, University of Toronto Law Journal 2010 = Bd. 60, S. 537–553.

De Angelis, Catherine, Jeffrey Drazen, Frank Frizelle, Charlotte Haug, John Hoey, Richard Horton, Sheldon Kotzin, Christine Laine, Ana Marusic, John Overbeke, Torben Schroeder, Hal Sox & Martin Van Der Weyden: Clinical Trial Registration: A Statement from the International Committee of Medical Journal Editors, New England Journal of Medicine 2004 = Bd. 351, S. 1250–1251.

De Long, Bradford & Kevin Lang: Are all Economic Hypotheses False?, Journal of Political Economy 1992 = Bd. 100, S. 1257–1272.

DeChurch, Leslie & Jessica Mesmer-Magnus: The cognitive underpinnings of effective teamwork: A meta-analysis, Journal of Applied Psychology 2010 = Bd. 95, S. 32–53.

DeChurch, Leslie & Jessica Mesmer-Magnus: Measuring shared team mental models: A meta-analysis, Group Dynamics: Theory, Research, and Practice 2010 = Bd. 14, S. 1–14.

Deci, Edward, Richard Koestner & Richard Ryan: A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation, Psychological Bulletin 1999 = Bd. 125, S. 627–668.

Del Mar, Chris, Paul Glasziou & Dan Mayer: Teaching Evidence Based Medicine, British Medical Journal 2004 = Bd. 329, S. 989–990.

Delarue, Anne, Geert Van Hootegem, Stephen Procter & Mark Burridge: Teamworking and organizational performance: A review of survey-based research, International Journal of Management Reviews 2008 = Bd. 10, S. 127–148.

DellaVigna, Stefano: Psychology and Economics: Evidence from the Field, Journal of Economic Literature 2009 = Bd. 47, S. 315–372.

Deutsch, Erwin: Rechtliche Grenzen des sozialpsychologischen Experiments, in: Ehmann/Hefermehl/Laufs, Privatautonomie, Eigentum und Verantwortung. Festgabe für Hermann Weitnauer zum 70. Geburtstag, Berlin 1980, S. 297–313.

Devers, Cynthia, Albert Cannella, Gregory Reilly & Michele Yoder: Executive Compensation: A Multidisciplinary Review of Recent Developments, Journal of Management 2007 = Bd. 33, S. 1016–1072.

Devine, Dennis, Laura Clayton, Benjamin Dunford, Rasmy Seying & Jennifer Pryce: Jury decision making: 45 years of empirical research on deliberating groups, Psychology, Public Policy, and Law 2001 = Bd. 7, S. 622–727.

Dewald, William, Jerry Thursby & Richard Anderson: Replication in Empirical Economics: The Journal of Money, Credit and Banking Project, The American Economic Review 1986 = Bd. 76, S. 587–603.

Diamond, Shari Seidman: Empirical Marine Life in Legal Waters: Clams, Dolphins, and Plankton, University of Illinois Law Review 2002, S. 803–818.

Dicks, Lynn: Bees, lies and evidence-based policy, Nature 2013 = Bd. 494, S. 283.

Diekmann, Andreas: Aktuelle Probleme der empirischen Sozialforschung, Kölner Zeitschrift für Soziologie und Sozialpsychologie. Sonderheft 44: Methoden der Sozialforschung 2004, S. 8–32.

Diekmann, Andreas: Are Most Published Research Findings False?, Jahrbücher für Nationalökonomie und Statistik 2011 = Bd. 231, S. 628–635.

Diekmann, Andreas: Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen, 23. Aufl. (6. Aufl. der Neuausg.), Reinbek 2012.

Dipboye, Robert & Michael Flanagan: Research settings in industrial and organizational psychology: Are findings in the field more generalizable than in the laboratory?, American Psychologist 1979 = Bd. 34, S. 141–150.

DNEbM: Wir über uns – Chronik des DNEbM, Deutsches Netzwerk Evidenzbasierte Medizin Zugriff am 13.12.2012 unter http://www.ebm-netzwerk.de/wer-wir-sind/chronik.

Dohmen, Thomas: Do professionals choke under pressure?, Journal of Economic Behavior & Organization 2008 = Bd. 65, S. 636–653.

Doralt, Peter, Christian Nowotny & Susanne Kalss (Hg.): Kommentar zum Aktiengesetz, 2 Bde., 2. Aufl., Wien 2012.

Dorff, Michael: The Group Dynamics Theory of Executive Compensation, Cardozo Law Review 2007 = Bd. 28, S. 2025–2081.

Dose, Stefan: Zivilrechtliche Haftung und Aufgabendelegation auf Ausschüsse im Aufsichtsrat der AG, Zeitschrift für Unternehmens- und Gesellschaftsrecht 1973, S. 300–316.

Dose, Stefan: Die Rechtsstellung der Vorstandsmitglieder einer Aktiengesellschaft, Diss. 1967, 3. Aufl., Köln 1975.

Doukas, John & Dimitris Petmezas: Acquisitions, Overconfident Managers and Self-attribution Bias, European Financial Management 2007 = Bd. 13, S. 531–577.

Doyen, Stéphane, Olivier Klein, Cora-Lise Pichon & Axel Cleeremans: Behavioral Priming: It's All in the Mind, but Whose Mind?, PLoS ONE 2012 e29081, Zugriff am 27.6.2014 unter http://dx.doi.org/10.1371%2Fjournal.pone.0029081.

Drewes, Sylvana, Thomas Schultze & Stefan Schulz-Hardt: Leistung in Gruppen, in: Frey/Bierhoff, Sozialpsychologie – Interaktion und Gruppe, Göttingen 2011, S. 221–244.

Dror, Itiel & Robert Rosenthal: Meta-analytically Quantifying the Reliability and Biasability of Forensic Experts, Journal of Forensic Sciences 2008 = Bd. 53, S. 900–903.

Dubben, Hans-Hermann & Hans-Peter Beck-Bornholdt: Die Bedeutung der statistischen Signifikanz, Kölner Zeitschrift für Soziologie und Sozialpsychologie. Sonderheft 44: Methoden der Sozialforschung 2004, S. 61–74.

Dubben, Hans-Hermann & Hans-Peter Beck-Bornholdt: Der Hund, der Eier legt. Erkennen von Fehlinformation durch Querdenken, 6. Aufl., Reinbek 2011.

Duckworth, Frank: The Royal Statistical Society Schools Lecture 2004: 'Lies and Statistics', Part 2, Teaching Statistics 2006 = Bd. 28, S. 84–89.

Dudenredaktion, Wissenschaftlicher Rat: Das große Fremdwörterbuch: Herkunft und Bedeutung der Fremdwörter, 2. Aufl., Mannheim 2000.

Dunning, David & Lee Ross: Overconfidence in Individual and Group Prediction: Is The Collective Any Wiser? (BEDR 90-02), Cornell University, Center for Behavioral Economics and Decision Research, Working Paper Series 1990, S. 3–45.

Durant, Robert, Robert Kramer, James Perry, Debra Mesch & Laurie Paarlberg: Motivating Employees in a New Governance Era: The Performance Paradigm Revisited, Public Administration Review 2006 = Bd. 66, S. 505–514.

Eberbach, Wolfram & Heinz Schuler: Zur Aufklärungspflicht bei psychologischen Experimenten, JuristenZeitung 1982, S. 356–363.

EBM Working Group: Evidence-based medicine: A new approach to teaching the practice of medicine, JAMA: The Journal of the American Medical Association 1992 = Bd. 268, S. 2420–2425.

ECO: 3d Session, 3d Parliament, 12 Victoria, 1849 Bill 14. An Act to provide for the formation of Incorporated Joint-Stock Companies, Early Canadiana Online, Zugriff am 4.10.2012 unter http://eco.canadiana.ca/view/oocihm.9_04695/4?r=0&s=1.

Economiesuisse: Swiss code of best practice for corporate governance, Economiesuisse – Verband der Schweizer Unternehmen, Zugriff am 27.6.2014 unter http://www.economiesuisse.ch/de/PDF Download Files/pospap_swiss-code_corp-govern_20080221_de.pdf.

Efron, Bradley: The 1977 Rietz Lecture: Bootstrap Methods: Another Look at the Jackknife, The Annals of Statistics 1979 = Bd. 7, S. 1–26.

Ehren, Thomas & Marius Gros: Anforderungen an eine angemessene Vorstandsvergütung aus Sicht der Praxis, Der Konzern. Zeitschrift für Gesellschaftsrecht, Steuerrecht, Bilanzrecht und Rechnungslegung der verbundenen Unternehmen 2010, S. 412–423.

Eidenmüller, Horst: Rechtswissenschaft als Realwissenschaft, JuristenZeitung 1999, S. 53–61.

Eidenmüller, Horst: Die GmbH im Wettbewerb der Rechtsformen, Zeitschrift für Unternehmens- und Gesellschaftsrecht 2007 = Bd. 36, S. 168–211.

Eidenmüller, Horst: Forschungsperspektiven im Unternehmensrecht, JuristenZeitung 2007 = Bd. 62, S. 487–494.

Eidenmüller, Horst: Forschungsperspektiven im Unternehmensrecht, Zeitschrift für Unternehmens- und Gesellschaftsrecht 2007 = Bd. 36, S. 484–499.

Eidenmüller, Horst: Liberaler Paternalismus, JuristenZeitung 2011 = Bd. 66, S. 814–821.

Eidenmüller, Horst: Effizienz als Rechtsprinzip. Möglichkeiten und Grenzen der ökonomischen Analyse des Rechts, Diss. 1994, 3. Aufl., Tübingen 2005.

Eidenmüller, Horst, Andreas Engert & Lars Hornuf: Die Societas Europaea: Empirische Bestandsaufnahme und Entwicklungslinien einer neuen Rechtsform, Die Aktiengesellschaft 2008, S. 721–730.

Eidenmüller, Horst, Andreas Engert & Lars Hornuf: Incorporating under European Law: The Societas Europaea as a Vehicle for Legal Arbitrage, European Business Organization Law Review 2009 = Bd. 10, S. 1–33.

Eidenmüller, Horst, Andreas Engert & Lars Hornuf: Vom Wert der Wahlfreiheit: Eine empirische Analyse der Societas Europaea als Rechtsformalternative, Die Aktiengesellschaft 2009, S. 845–855.

Eidenmüller, Horst, Andreas Engert & Lars Hornuf: How Does the Market React to the Societas Europaea?, European Business Organization Law Review 2010 = Bd. 11, S. 35–50.

Eisele, Florian & Andreas Walter: Kursreaktionen auf die Ankündigung von Going Private-Transaktionen am deutschen Kapitalmarkt, Zeitschrift für betriebswirtschaftliche Forschung 2006 = Bd. 58, S. 337–362.

Eisenberg, Theodore: Empirical Methods and the Law, Journal of the American Statistical Association 2000 = Bd. 95, S. 665–669.

Eisenberg, Theodore: Why Do Empirical Legal Scholarship?, San Diego Law Review 2004 = Bd. 41, S. 1741–1746.

Eisenberg, Theodore: The Origins, Nature, and Promise of Empirical Legal Studies and a Response to Concerns, University of Illinois Law Review 2011, S. 1713–1738.

Elliott, Donald Law and Biology: The New Synthesis, Saint Louis University Law Journal 1997 = Bd. 41, S. 595–624.

Ellis, Paul: The Essential Guide to Effect Sizes: Statistical Power, Meta-Analysis, and the Interpretation of Research Results, Cambridge 2010.

Emde, Ernst Thomas: Gesamtverantwortung und Ressortverantwortung im Vorstand der AG, in: Burgard/Hadding/Mülbert/Nietsch/Weiter, Festschrift für Uwe H. Schneider zum 70. Geburtstag, Köln 2011, S. 295–321.

Engel, Christoph: Rechtswissenschaft als angewandte Sozialwissenschaft. Die Aufgabe der Rechtswissenschaft nach der Öffnung der Rechtsordnung für sozialwissenschaftliche Theorie, in: Engel, Methodische Zugänge zu einem Recht der Gemeinschaftsgüter, Baden-Baden 1998, S. 11–40.

Engel, Christoph: Verhaltenswissenschaftliche Analyse: eine Gebrauchsanweisung für Juristen, in: Engel/Englerth/Lüdemann/Spiecker gen. Döhmann, Recht und Verhalten. Beiträge zu Behavioral Law and Economics, Tübingen 2007, S. 363–405.

Engel, Christoph: The behaviour of corporate actors: How much can we learn from the experimental literature?, Journal of Institutional Economics 2010 = Bd. 6, S. 445–475.

Engel, Christoph: The Multiple Uses of Experimental Evidence in Legal Scholarship, Journal of Institutional and Theoretical Economics 2010 = Bd. 166, S. 199–202.

Engel, Christoph, Markus Englerth, Jörn Lüdemann & Indra Spiecker gen. Döhmann: Recht und Verhalten. Beiträge zu Behavioral Law and Economics, Tübingen 2007.

Engel, Christoph & Wolfgang Schön: Vorwort, in: Engel/Schön, Das Proprium der Rechtswissenschaft, Tübingen 2007, S. IX–XIV.

Engisch, Karl: Logische Studien zur Gesetzesanwendung, 3. Aufl., Heidelberg 1963.

Englerth, Markus: Behavioral Law and Economics − eine kritische Einführung, in: Engel/Englerth/Lüdemann/Spiecker gen. Döhmann, Recht und Verhalten. Beiträge zu Behavioral Law and Economics, Tübingen 2007, S. 60–130.

Englerth, Markus: Der beschränkt rationale Verbrecher. Behavioral Economics in der Kriminologie, Diss. 2010, Berlin/Münster 2010.

Enserink, Martin The Joy of Evidence-Based Cooking, Science 2006 = Bd. 314, S. 1235–1236.

Epstein, Lee & Gary King: A Defense of Empirical Legal Scholarship: A Reply, The University of Chicago Law Review 2002 = Bd. 69, S. 191–209.

Epstein, Lee & Gary King: The Rules of Inference, The University of Chicago Law Review 2002 = Bd. 69, S. 1–133.

Epstein, Lee & Andrew Martin: Quantitative Approaches to Empirical Legal Research, in: Cane/Kritzer, The Oxford Handbook of Empirical Legal Research, Oxford 2010, S. 901–925.

Epstein, Lee, Andrew Martin & Christina Boyd: On the Effective Communication of the Results of Empirical Studies, Part II, Vanderbilt Law Review 2007 = Bd. 60, S. 801–846.

Epstein, Lee, Andrew Martin & Matthew Schneider: On the Effective Communication of the Results of Empirical Studies, Part I, Vanderbilt Law Review 2006 = Bd. 59, S. 1811–1871.

Epstein, Richard: Let “The Fundamental Things Apply”: Necessary and Contingent Truths in Legal Scholarship, Harvard Law Review 2002 = Bd. 115, S. 1288–1313.

Erle, Bernd: Das Vetorecht des Vorstandsvorsitzenden in der AG, Die Aktiengesellschaft 1987, S. 7–12.

ETH-KOF: Umfragen: Strukturumfragen, ETH Zürich Konjunkturforschungsstelle, Zugriff am 14.10.2012 unter http://www.kof.ethz.ch/de/umfragen/strukturumfragen/.

Europäische Kommission: Vorschlag einer Verordnung (EWG) des Rates über das Statut für europäische Aktiengesellschaften, EUR-Lex, Zugriff am 3.2.2013 unter http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:1970:124:FULL:DE:PDF.

Europäische Kommission: Proposal for a Council regulation on the Statute for European companies, COM (75) 150 final, Bulletin of the European Communities (University of Pittsburgh: Archive of European Integration), Zugriff am 3.2.2013 unter http://aei.pitt.edu/8558/1/8558.pdf.

Europäische Kommission: Geänderter Vorschlag für eine Verordnung (EWG) des Rates über das Statut der Europäischen Aktiengesellschaft, EUR-Lex: KOM(91) 174 endg. – SYN 218, Zugriff am 8.4.2012 unter http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:1991:176:0001:0068:DE:PDF.

Europäische Kommission: Commission Staff Working Document, Corporate Governance in Financial Institutions: Lessons to be drawn from the current financial crisis, best practices, SEC (2010) 669, Zugriff am 15.4.2012 unter http://ec.europa.eu/internal_market/company/docs/modern/sec2010_669_en.pdf.

Europäische Kommission: Grünbuch Europäischer Corporate Governance-Rahmen, KOM (2011) 164/3, Zugriff am 8.4.2012 unter http://ec.europa.eu/internal_market/company/docs/modern/com2011-164_de.pdf.

Fama, Eugene, Lawrence Fisher, Michael Jensen & Richard Roll: The Adjustment of Stock Prices to New Information, International Economic Review 1969 = Bd. 10, S. 1–21.

Fama, Eugene & Michael Jensen: Separation of Ownership and Control, Journal of Law & Economics 1983 = Bd. 26, S. 301–325.

Fanto, James: Braking the Merger Momentum: Reforming Corporate Law Governing Mega-Mergers, Buffalo Law Review 2001 = Bd. 49, S. 249–358.

Fanto, James: Quasi-Rationality in Action: A Study Of Psychological Factors in Merger Decision-Making, Ohio State Law Journal 2001 = Bd. 62, S. 1333–1407.

Fanto, James: Whistleblowing and the Public Director: Countering Corporate Inner Circles, Oregon Law Review 2004 = Bd. 83, S. 435–539.

Farley, John, Donald Lehmann & Lane Mann: Designing the Next Study for Maximum Impact, Journal of Marketing Research 1998 = Bd. 35, S. 496–501.

Feinstein, Alvan & Ralph Horwitz: Problems in the ‘‘Evidence’’ of ‘‘Evidence-based Medicine’’, The American Journal of Medicine 1997 = Bd. 103, S. 529–535.

Feltovich, Nick: What’s to Know About Laboratory Experimentation in Economics?, Journal of Economic Surveys 2011 = Bd. 25, S. 371–379.

Ferraro, Fabrizio, Jeffrey Pfeffer & Robert Sutton: Economics Language and Assumptions: How Theories Can Become Self-Fulfilling, Academy of Management Review 2005 = Bd. 30, S. 8–24.

Feuerbach, Anselm von: Betrachtungen über die Oeffentlichkeit und Mündlichkeit der Gerechtigkeitspflege, Gießen 1821.

Finch, Janet: The Vignette Technique in Survey Research, Sociology 1987 = Bd. 21, S. 105–114.

Finkelstein, Michael: Basic Concepts of Probability and Statistics in the Law, New York 2009.

Fiorina, Morris & Charles Plott: Committee Decisions under Majority Rule: An Experimental Study, The American Political Science Review 1978 = Bd. 72, S. 575–598.

Fisch, Gene: Evaluating data from behavioral analysis: visual inspection or statistical models?, Behavioural Processes 2001 = Bd. 54, S. 137–154.

Fischer, Sven, Sebastian Goerg & Hanjo Hamann: Cui Bono, Benefit Corporation? An Experiment Inspired by Social Enterprise Legislation in Germany and the US, Social Science Research Network, Zugriff am 3.8.2014 unter http://ssrn.com/abstract=2106190.

Fleischer, Holger: Die „Business Judgment Rule“ im Spiegel von Rechtsvergleichung und Rechtsökonomie, in: Wank/Hirte/Frey/Fleischer/Thüsing, Festschrift für Herbert Wiedemann zum 70. Geburtstag, München 2002, S. 827–849.

Fleischer, Holger: Zum Grundsatz der Gesamtverantwortung im Aktienrecht, Neue Zeitschrift für Gesellschaftsrecht 2003, S. 449–459.

Fleischer, Holger: *Behavioral Law and Economics* im Gesellschafts- und Kapitalmarktrecht – ein Werkstattbericht, in: Fuchs/Schwintowski/Zimmer, Wirtschafts- und Privatrecht im Spannungsfeld von Privatautonomie, Wettbewerb und Regulierung. Festschrift für Ulrich Immenga zum 70. Geburtstag, München 2004, S. 575–587.

Fleischer, Holger: Gesellschafts- und Kapitalmarktrecht als wissenschaftliche Disziplin. Das Proprium der Rechtswissenschaft, in: Engel/Schön, Das Proprium der Rechtswissenschaft, Tübingen 2007, S. 50–76.

Fleischer, Holger: Finanzinvestoren im ordnungspolitischen Gesamtgefüge von Aktien-, Bankaufsichts- und Kapitalmarktrecht, Zeitschrift für Unternehmens- und Gesellschaftsrecht 2008, S. 185–224.

Fleischer, Holger: Zukunftsfragen der Corporate Governance in Deutschland und Europa: Aufsichtsräte, institutionelle Investoren, Proxy Advisors und Whistleblowers, Zeitschrift für Unternehmens- und Gesellschaftsrecht 2011, S. 155–181.

Fleischer, Holger & Wulf Goette (Hg.): Münchener Kommentar zum GmbHG 3Bde., 1. Aufl., München 2011.

Fleischer, Holger & Daniel Zimmer: Effizienz als Regelungsziel im Handels- und Wirtschaftsrecht, Frankfurt/M. 2008.

Flick, Uwe: Qualitative Sozialforschung: Eine Einführung, 3. Aufl., Reinbek 2010.

Florence, Francisco: Das brasilianische Gesellschaftsrecht, 4. Aufl., Baden-Baden 2005.

Fonk, Hans-Joachim: Die Zulässigkeit von Vorstandsbezügen dem Grunde nach. Aktienrechtliche Anmerkungen zum Urteil des LG Düsseldorf, NZG 2004, 1057 – Mannesmann, Neue Zeitschrift für Gesellschaftsrecht 2005, S. 248–254.

Forbes, Daniel & Frances Milliken: Cognition and Corporate Governance: Understanding Boards of Directors as Strategic Decision-Making Groups, The Academy of Management Review 1999 = Bd. 24, S. 489–505.

Forstmoser, Peter: Monistische oder dualistische Unternehmensverfassung? Das Schweizer Konzept, Zeitschrift für Unternehmens- und Gesellschaftsrecht 2003, S. 688–719.

Fraidin, Samuel: Duty of Care Jurisprudence: Comparing Judicial Intuition and Social Psychology Research, U.C. Davis Law Review 2004 = Bd. 38, S. 1–78.

Fraidin, Samuel: When is one head better than two? Interdependent information in group decision making, Organizational Behavior and Human Decision Processes 2004 = Bd. 93, S. 102–113.

FRC: The UK Corporate Governance Code, Financial Reporting Council, Zugriff am 10.10.2012 unter http://www.frc.org.uk/getattachment/a7f0aa3a-57dd-4341-b3e8-ffa99899e154/UK-Corporate-Governance-Code-September-2012.aspx.

Frey, Kaspar: Gesellschaftsrecht als Systembaukasten, Neue Zeitschrift für Gesellschaftsrecht 2004, S. 169–177.

Frooman, Jeff: Socially Irresponsible and Illegal Behavior and Shareholder Wealth, Business & Society 1997 = Bd. 36, S. 221–249.

Fuess, Scott: On Replication in Business and Economics Research: The QJBE Case, Quarterly Journal of Business and Economics 1996 = Bd. 35, S. 3–13.

Gangl, Markus & Thomas DiPrete: Kausalanalyse durch Matchingverfahren, Kölner Zeitschrift für Soziologie und Sozialpsychologie. Sonderheft 44: Methoden der Sozialforschung 2004, S. 396–420.

Gardner, Donald, Linn Van Dyne & Jon Pierce: The effects of pay level on organization-based self-esteem and performance: A field study, Journal of Occupational and Organizational Psychology 2004 = Bd. 77, S. 307–322.

Garrett, Ray: John Doe Incorporates Himself, The Business Lawyer 1964 = Bd. 19, S. 535–537.

Garrn, Heino: Zur Rationalität rechtlicher Entscheidungen, Stuttgart 1986.

Geiler, Karl: Die wirtschaftsrechtliche Methode im Gesellschaftsrecht, in: Predari/Schlegelberger/Wolff, Beiträge zur Erläuterung des Deutschen Rechts, Berlin 1927, S. 593–619.

Gelman, Andrew, Cristian Pasarica & Rahul Dodhia: Let's Practice What We Preach: Turning Tables into Graphs, The American Statistician 2002 = Bd. 56, S. 121–130.

Gensler, Sonja, Bernd Skiera & Martin Böhm: Einsatzmöglichkeiten der Matching Methode zur Berücksichtigung von Selbstselektion, Journal für Betriebswirtschaft 2005 = Bd. 55, S. 37–62.

Gerling, Kerstin, Hans Peter Grüner, Alexandra Kiel & Elisabeth Schulte: Information acquisition and decision making in committees: A survey, European Journal of Political Economy 2005 = Bd. 21, S. 563–597.

Gerpott, Torsten & Nejc Jakopin: Ereignisstudien: eine kapitalmarktorientierte Methodik zur Bewertung von neuen unternehmensbezogenen Informationen, Wirtschaftswissenschaftliches Studium. Zeitschrift für Studium und Forschung 2006 = Bd. 35, S. 66–72.

Gerum, Elmar: Das deutsche Corporate Governance-System. Eine empirische Untersuchung, Stuttgart 2007.

Gessner, Volkmar: Probleme der Zusammenarbeit zwischen Juristen und Sozialwissenschaftlern. Erfahrungen und Folgerungen aus einem Informationsbesuch nordamerikanischer Universitäten, JuristenZeitung 1971, S. 324–327.

Gessner, Volkmar: Rechtssoziologie und Rechtspraxis. Zur Rezeption empirischer Rechtsforschung, in: Plett/Ziegert, Empirische Rechtsforschung zwischen Wissenschaft und Politik. Zur Problemlage rechtssoziologischer Auftragsforschung, Tübingen 1984, S. 69–112.

Gessner, Volkmar: Rechtssoziologie und Rechtspraxis: Zur Rezeption empirischer Rechtsforschung, Soziale Welt 1984 = Bd. 35, S. 480–501.

Getman, Julius: Contributions of Empirical Data to Legal Research, Journal of Legal Education 1985 = Bd. 35, S. 489–494.

Gevurtz, Franklin: The Historical and Political Origins of the Corporate Board of Directors, Hofstra Law Review 2004 = Bd. 33, S. 89–173.

Gigerenzer, Gerd: Mindless statistics, The Journal of Socio-Economics 2004 = Bd. 33, S. 587–606.

Gigone, Daniel & Reid Hastie: The common knowledge effect: Information sharing and group judgment, Journal of Personality and Social Psychology 1993 = Bd. 65, S. 959–974.

Gigone, Daniel & Reid Hastie: Proper analysis of the accuracy of group judgments, Psychological Bulletin 1997 = Bd. 121, S. 149–167.

Gillan, Stuart: Recent Developments in Corporate Governance: An Overview, Journal of Corporate Finance 2006 = Bd. 12, S. 381–402.

Gilles, Peter: Rechtstatsachenforschung und Rechtsstaat. Ein Beitrag zur gegenwärtigen „Krise des Rechtsstaats“ und zum gegenwärtigen „Dilemma der Rechtstatsachenforschung“ im Anschluss an ein deutsch-türkisches Symposium in Istanbul 1994, in: Prütting/Rüssmann, Verfahrenssrecht am Ausgang des 20. Jahrhunderts. Festschrift für Gerhard Lüke zum 70. Geburtstag, München 1997, S. 139–165.

Ginsburg, Tom & Thomas Miles: Empiricism and the Rising Incidence of Coauthorship in Law, University of Illinois Law Review 2011, S. 1785–1825.

Girvan, Erik & Grace Deason: Social Science in Law: A Psychological Case For Abandoning the “Discriminatory Motive” under Title VII, Cleveland State Law Review 2013 = Bd. 60, S. 1057–1101.

Gist, Marilyn, Edwin Locke & Susan Taylor: Organizational Behavior: Group Structure, Process, and Effectiveness, Journal of Management 1987 = Bd. 13, S. 237–257.

Glaser, Markus, Philipp Schäfers & Martin Weber: Managerial Optimism and Corporate Investment: Is the CEO Alone Responsible for the Relation?, Social Science Research Network, Zugriff am 15.3.2012 unter http://ssrn.com/paper=967649.

Gleick, James: Chaos – die Ordnung des Universums. Vorstoß in Grenzbereiche der modernen Physik, 6. Aufl., München 1994.

Goette, Wulf & Mathias Habersack (Hg.): Münchener Kommentar zum Aktiengesetz Bd. 7: Europäisches Aktienrecht, 3. Aufl., München 2012.

Goette, Wulf & Mathias Habersack (Hg.): Münchener Kommentar zum Aktiengesetz Bd. 2: §§ 76–117, MitbestG, DrittelbG, 4. Aufl., München 2014.

Goldacre, Ben: Backwards step on looking into the future, The Guardian, Zugriff am 28.6.2012 unter http://www.guardian.co.uk/commentisfree/2011/apr/23/ben-goldacre-bad-science.

Goldsmith, Jack & Adrian Vermeule: Empirical Methodology and Legal Scholarship, The University of Chicago Law Review 2002 = Bd. 69, S. 153–167.

Gonick, Larry & Woollcott Smith: Statistik in Cartoons, München 2009.

Good, Irving John & Y. Mittal: The Amalgamation and Geometry of Two-by-Two Contingency Tables, The Annals of Statistics 1987 = Bd. 15, S. 694–711.

Good, Phillip & James Hardin: Common Errors in Statistics (and How to Avoid Them), 3. Aufl., Hoboken 2009.

Gorard, Stephen: Towards a Judgement-Based Statistical Analysis, British Journal of Sociology of Education 2006 = Bd. 27, S. 67–80.

Gosling, Tom: Making executive pay work: The psychology of incentives, PwC United Kingdom, Zugriff am 4.9.2012 unter http://www.pwc.co.uk/human-resource-services/publications/making-executive-pay-work.jhtml.

Gov't of India: Companies Act, Ministry of Corporate Affairs, Zugriff am 16.3.2014 unter http://www.mca.gov.in/Ministry/companies_act.html.

Graumann, Matthias: Der Entscheidungsbegriff in § 93 Abs. 1 Satz 2 AktG – Rekonstruktion des traditionellen Verständnisses und Vorschlag für eine moderne Konzeption, Zeitschrift für Unternehmens- und Gesellschaftsrecht 2011 = Bd. 40, S. 293–303.

Gravenstein, Viky: Französisches Gesellschaftsrecht, 3. Aufl., Frankfurt/M. 1988.

Grechenig, Kristoffel: Spanisches Aktien- und GmbH-Recht: das einstufige Verwaltungssystem in Beziehung zur Hauptversammlung und zu Gesellschafterrechten, Diss. 2004, Wien 2005.

Grechenig, Kristoffel & Martin Gelter: Divergente Evolution des Rechtsdenkens – Von amerikanischer Rechtsökonomie und deutscher Dogmatik, Rabels Zeitschrift für Ausländisches und Internationales Privatrecht 2008 = Bd. 72, S. 513–561.

Grechenig, Kristoffel & Martin Gelter: The Transatlantic Divergence in Legal Thought: American Law and Economics vs. German Doctrinalism, Hastings International and Comparative Law Review 2008 = Bd. 31, S. 295–360.

Green, Christopher: Classics in the History of Psychology, York University, Toronto, Canada, Zugriff am 16.7.2012 unter http://psychclassics.yorku.ca/author.htm.

Green, Judith & Nicky Britten: Qualitative research and evidence based medicine, British Medical Journal 1998 = Bd. 316, S. 1230–1232.

Green, Samuel: How Many Subjects Does It Take To Do A Regression Analysis, Multivariate Behavioral Research 1991 = Bd. 26, S. 499–510.

Greenhalgh, Trisha: Narrative based medicine in an evidence based world, British Medical Journal 1999 = Bd. 318, S. 323–325.

Greenwald, Anthony: Consequences of Prejudice Against the Null Hypothesis, Psychological Bulletin 1975 = Bd. 82, S. 1–20.

Grigoleit, Jens: Kapitalmarktreaktionen auf die Ankündigung des Wechsels von Vorstandsvorsitzenden in den Aufsichtsrat bei deutschen Unternehmen, Zeitschrift für Planung & Unternehmenssteuerung 2011 = Bd. 21, S. 131–157.

Grissom, Robert: The magical number .7±.2: Meta-meta-analysis of the probability of superior outcome in comparisons involving therapy, placebo, and control, Journal of Consulting and Clinical Psychology 1996 = Bd. 64, S. 973–982.

Gröls, Marcel & Tanja Gröls: Ein Ranking juristischer Fachzeitschriften, JuristenZeitung 2009, S. Beil. H. 17, 32–43.

Groß, Jochen & Christina Börensen: Wie valide sind Verhaltensmessungen mittels Vignetten? Ein methodischer Vergleich von faktoriellem Survey und Verhaltensbe-obachtung, in: Kriwy/Gross, Klein aber fein! Quantitative empirische Sozialforschung mit kleinen Fallzahlen, Wiesbaden 2009, S. 149–178.

Groß, Thomas: Das Kollegialprinzip in der Verwaltungsorganisation, Habil. 1998, Tübingen 1999.

Grundei, Jens: Entscheidungsfindung im Vorstand deutscher Aktiengesellschaften: Organisationstheoretische Anmerkungen zu möglichen Diskrepanzen zwischen Recht und Wirklichkeit, Diskussionspapier 01/2004 des Berlin Center of Corporate Governance, Zugriff am unter

Grundei, Jens & Axel von Werder: Die Angemessenheit der Informationsgrundlage als Anwendungsvoraussetzung der Business Judgment Rule – Anforderungen an die Fundierung strategischer Entscheidungen aus betriebswirtschaftlicher Sicht, Die Aktiengesellschaft 2005, S. 825–834.

Gruter, Margaret: Die Bedeutung der Verhaltensforschung für die Rechtswissenschaft, Berlin 1976.

Gruter, Margaret: Rechtsverhalten. Biologische Grundlagen mit Beispielen aus dem Familien- und Umweltrecht, Köln 1993.

Gu, Minkang: Understanding Chinese Company Law, Hong Kong 2006.

Guillard, Dorothee: Einzelrichter und Kollegialgericht seit der Reichszivilprozeßordnung von 1877 bis zur Zivilprozeßnovelle von 1933, Diss. 1989, Tübingen 1989.

Gully, Stanley, Kara Incalcaterra, Aparna Joshi & Matthew Beaubien: A meta-analysis of team-efficacy, potency, and performance: Interdependence and level of analysis as moderators of observed relationships, Journal of Applied Psychology 2002 = Bd. 87, S. 819–832.

Gusy, Christoph: „Wirklichkeit“ in der Rechtsdogmatik, JuristenZeitung 1991 = Bd. 46, S. 213–222.

Güttler, André: Wird die Ad-hoc-Publizität korrekt umgesetzt? Eine empirische Analyse unter Einbezug von Unternehmen des Neuen Markts, Zeitschrift für betriebswirtschaftliche Forschung 2005 = Bd. 57, S. 237–259.

Guttman, Louis: The illogic of statistical inference for cumulative science, Applied Stochastic Models and Data Analysis 1985 = Bd. 1, S. 3–9.

Guyatt, Gordon, Deborah Cook & Brian Haynes: Evidence based medicine has come a long way, British Medical Journal 2004 = Bd. 329, S. 990–991.

Guzzo, Richard & Marcus Dickson: Teams in Organizations: Recent Research on Performance and Effectiveness, Annual Review of Psychology 1996 = Bd. 47, S. 307–338.

Habersack, Mathias (Hg.): Münchener Kommentar zum Bürgerlichen Gesetzbuch Bd. 5: §§ 705-853 BGB, PartGG, ProdHaftG, 6. Aufl., München 2013.

Habscheid, Walther: Urteilskritik durch am Verfahren „beteiligte“ Rechtsanwälte, Professoren und Richter, Neue Juristische Wochenschrift 1999, S. 2230–2233.

Häcker, Hartmut: Die experimentelle Methode in der Rechtstatsachenforschung: Empirische Untersuchungen zum Problem Einzelrichter und Kollegium, in: Bender, Tatsachenforschung in der Justiz. Einführung in die Methoden, Tübingen 1972, S. 143–166.

Hackman, Richard & Charles Morris: Group Tasks, Group Interaction Process, and Group Performance Effectiveness: A Review and Proposed Integration, Advances in Experimental Social Psychology 1975 = Bd. 8, S. 45–99.

Haft, Fritjof, Hagen Hof & Steffen Wesche: Bausteine zu einer Verhaltenstheorie des Rechts, Baden-Baden 2001.

Haft, Robert: Business Decisions by the New Board: Behavioral Science and Corporate Law, Michigan Law Review 1981 = Bd. 80, S. 1–67.

Hagen, Johann: Statistik für Juristen, Wien 2005.

Hall, Kath: Looking beneath the surface. The impact of psychology on corporate decision making, Managerial Law 2007 = Bd. 49, S. 93–105.

Ham, Chris, David Hunter & Ray Robinson: Evidence based policymaking, British Medical Journal 1995 = Bd. 310, S. 71–72.

Hamann, Andreas: Das Kollegialprinzip und der „Einzelrichter" nach der 4. Novelle zur VwGO, Verwaltungsarchiv. Zeitschrift für Verwaltungslehre, Verwaltungsrecht und Verwaltungspolitik 1992 = Bd. 83, S. 201–212.

Hamann, Hanjo: Was kostet ein Manager? Preisfindungsmechanismen bei der Vorstandsvergütung, Bonner Rechtsjournal 2010, S. 27–32.

Hamann, Hanjo: Der „Sprachgebrauch" im Waffenarsenal der Jurisprudenz. Die Rechtspraxis im Spiegel der quantitativ-empirischen Sprachforschung, in: Vogel, Zugänge zur Rechtssemantik. Interdisziplinäre Ansätze im Zeitalter neuer Medien, Berlin 2014, i.E.

Hamann, Wolfgang: „Iudex non calculat" – Recht und Statistik, in: Schröder/Clausen/Behr, Essener Beiträge zur empirischen Wirtschaftsforschung. Festschrift für Prof. Dr. Walter Assenmacher, Wiesbaden 2012, S. 307–316.

Hamblin, Terence John: Fake!, British Medical Journal 1981 = Bd. 283, S. 1671–1674.

Hanau, Peter: Zur Zuständigkeit des Arbeitsdirektors (§ 33 MitbestG) für leitende Angestellte und Unternehmenssparten, Zeitschrift für Unternehmens- und Gesellschaftsrecht 1983 = Bd. 12, S. 346–375.

Hansen, Herbert: Die Aktiengesellschaften im Jahre 1970. Eine Auswertung statistischer Unterlagen, Die Aktiengesellschaft 1971, S. 148–149.

Hansen, Herbert: Die Entwicklung der GmbH, GmbH-Rundschau 1977, S. 145–146.

Hansen, Herbert: Der gestiegene wirtschaftliche Stellenwert der GmbH, GmbH-Rundschau 2004, S. 39–42.

Harper, Douglas: Online Etymology Dictionary, Zugriff zuletzt am 27.6.2014 unter http://www.etymonline.com.

Harrison, Glenn & John List: Field Experiments, Journal of Economic Literature 2004 = Bd. 42, S. 1009–1055.

Hart, Oliver: Corporate Governance: Some Theory and Implications, The Economic Journal 1995 = Bd. 105, S. 678–689.

Hassemer, Winfried: Zum Spannungsverhältnis von Freiheit und Sicherheit. Drei Thesen, Vorgänge: Zeitschrift für Bürgerrechte und Gesellschaftspolitik 2002, S. 3/10–15.

Häusermann, Daniel: Fehler erster und zweiter Art im Strafrecht – oder was das Nichtingangsetzen der Parkuhr mit induktiver Statistik zu tun hat, Schweizerische Zeitschrift für Strafrecht 2011 = Bd. 2, S. 194–210.

Hayes, Brian: Ein Abenteuer in *n* Dimensionen, Spektrum der Wissenschaft 2012, S. 52–58.

Haymann, Franz: Die Mehrheitsentscheidung. Ihr Sinn und ihre Schranke, in: Tatarin-Tarnheyden, Festgabe für Rudolf Stammler zum 70. Geburtstage, Berlin 1926, S. 395–476.

Hayward, Mathew & Donald Hambrick: Explaining the Premiums Paid for Large Acquisitions: Evidence of CEO Hubris, Administrative Science Quarterly 1997 = Bd. 42, S. 103–127.

Heath, Chip & Forest Jourden: Illusion, Disillusion, and the Buffering Effect of Groups, Organizational Behavior and Human Decision Processes 1997 = Bd. 69, S. 103–116.

Heath, Chip, Richard Larrick & Joshua Klayman: Cognitive repairs: How organizational practices can compensate for individual shortcomings, Research in Organizational Behavior 1998 = Bd. 20, S. 1–37.

Heaton, James: Managerial Optimism and Corporate Finance, Financial Management 2002 = Bd. 31, S. 33–45.

Hedges, Larry & Ingram Olkin: Vote-counting methods in research synthesis, Psychological Bulletin 1980 = Bd. 88, S. 359–369.

Heermann, Peter: Beschränkung der persönlichen Haftung des Vereinsvorstands durch Ressortverteilung, in: Crezelius/Hirte/Vieweg, Festschrift für Volker Röhricht zum 65. Geburtstag – Gesellschaftsrecht, Rechnungslegung, Sportrecht, Köln 2005, S. 1191–1203.

Heinicke, Christiane: Local Lorentz Force Velocimetry for Liquid Metal Duct Flows, Diss., Ilmenau 2013. Zugriff am 3.8.2014 unter http://www.db-thueringen.de/servlets/DocumentServlet?id=22402.

Heinze, Christian: Vom Publizieren in der Sache „des Mandanten“. Ein Betrag zur Sittenbildung, Betriebs-Berater 1984, S. 502–504.

Heise, Michael: The Importance of Being Empirical, Pepperdine Law Review 1999 = Bd. 26, S. 807–834.

Heise, Michael: An Empirical Analysis of Empirical Legal Scholarship Production, 1990–2009, University of Illinois Law Review 2011, S. 1739–1752.

Heldrich, Andreas: Die Bedeutung der Rechtssoziologie für das Zivilrecht, Archiv für die civilistische Praxis 1986 = Bd. 186, S. 74–114.

Hellgardt, Alexander: Fehlerhafte Ad-hoc-Publizität als strafbare Marktmanipulation. Der Beweis von Taterfolg und Kausalität, Zeitschrift für Wirtschaftsrecht 2005, S. 2000–2008.

Helmis, Sven: Der Ausschluss von Minderheitsaktionären. Empirische Erfahrungen der ersten zehn Monate nach Inkrafttreten der neuen Regelung, Zeitschrift für Bankrecht und Bankwirtschaft 2003, S. 161–176.

Hendel, Dieter: Einzelrichter – oder Kollegialsystem. Hypothesen der Kleingruppenforschung, in: Bender, Tatsachenforschung in der Justiz. Einführung in die Methoden, Tübingen 1972, S. 105–125.

Hendel, Dieter: Hypothesen der Kleingruppenforschung zum Thema „Einzelrichter oder Kollegium“, JuristenZeitung 1973, S. 412–415.

Henrich, Joseph, Steven Heine & Ara Norenzayan: Most people are not WEIRD, Nature 2010 = Bd. 466, S. 29.

Henrich, Joseph, Steven Heine & Ara Norenzayan: The weirdest people in the world?, Behavioral and Brain Sciences 2010 = Bd. 33, S. 61–83.

Henry, P. J.: College Sophomores in the Laboratory Redux: Influences of a Narrow Data Base on Social Psychology's View of the Nature of Prejudice, Psychological Inquiry 2008 = Bd. 19, S. 49–71.

Henssler, Martin & Lutz Strohn (Hg.): Gesellschaftsrecht (Kurz-Kommentar), 2. Aufl., München 2014.

Herberger, Maximilian & Dieter Simon: Wissenschaftstheorie für Juristen. Logik, Semiotik, Erfahrungswissenschaften, Universität des Saarlandes. Institut für Rechtsinformatik, Archiv, Zugriff am 19.12.2012 unter http://archiv.jura.uni-saarland.de/wtheorie/Buch_Wissenschaftstheorie.pdf.

Hermalin, Benjamin & Michael Weisbach: The Effects of Board Composition and Direct Incentives on Firm Performance, Financial Management 1991 = Bd. 20, S. 101–112.

Hermes, Judith: Wunsch und Wirklichkeit der GmbH-Haftungsbeschränkung. Eine Rechtstatsachenuntersuchung, Diss. 2005, Baden-Baden 2006.

Hertwig, Ralph & Andreas Ortmann: Experimental practices in economics: A methodological challenge for psychologists?, Behavioral and Brain Sciences 2001 = Bd. 24, S. 383–403.

Heukelom, Floris: How validity travelled to economic experimenting, Journal of Economic Methodology 2011 = Bd. 18, S. 13–28.

Hewison, Alistair: Evidence-based medicine: what about evidence-based management?, Journal of Nursing Management 1997 = Bd. 5, S. 195–198.

Higgins, Julian & Sally Green: Cochrane Handbook for Systematic Reviews of Interventions, The Cochrane Collaboration, Zugriff am 20.12.2012 unter http://handbook.cochrane.org/.

Hill, Denise, Sheldon Hanton, Nic Matthews & Scott Fleming: Choking in sport: a review, International Review of Sport and Exercise Psychology 2010 = Bd. 3, S. 24–39.

Hill, Gayle: Group versus individual performance: Are N+1 heads better than one?, Psychological Bulletin 1982 = Bd. 91, S. 517–539.

Ho, Daniel & Donald Rubin: Credible Causal Inference for Empirical Legal Studies, Annual Review of Law and Social Science 2011 = Bd. 7, S. 17–40.

Höcker, Ralf: Lexikon der Rechtsirrtümer. Zechprellerei, Beamtenbeleidigung und andere juristische Volksmythen, 12. Aufl., Berlin 2007.

Hof, Hagen: Rechtsethologie. Recht im Kontext von Verhalten und außerrechtlicher Verhaltensregelung, Heidelberg 1996.

Hof, Hagen: Soziologie, Rechtssoziologie und Rechtsethologie. Gemeinsamkeiten, Unterschiede und neue Aufgaben, in: Brand/Strempel, Soziologie des Rechts. Festschrift für Erhard Blankenburg zum 60. Geburtstag, Baden-Baden 1998, S. 65–77.

Hoffmann-Becking, Michael: Zur rechtlichen Organisation der Zusammenarbeit im Vorstand der AG, Zeitschrift für Unternehmens- und Gesellschaftsrecht 1998, S. 497–519.

Hoffmann-Becking, Michael: Vorstandsvorsitzender oder CEO?, Neue Zeitschrift für Gesellschaftsrecht 2003, S. 745–750.

Hoffmann-Becking, Michael: Rechtliche Anmerkungen zur Vorstands- und Aufsichtsratsvergütung, Zeitschrift für das gesamte Handels- und Wirtschaftsrecht 2005 = Bd. 169, S. 155–180.

Hoffmann-Becking, Michael & Peter Rawert (Hg.): Beck'sches Formularbuch Bürgerliches, Handels- und Wirtschaftsrecht 10. Aufl., München 2010.

Hogarth, Robin: Discussion of An Assessment of Laboratory Experiments in Accounting, Journal of Accounting Research 1982 = Bd. 20, S. S108–S116.

Holland, Paul: Statistics and Causal Inference, Journal of the American Statistical Association 1986 = Bd. 81, S. 945–970.

Holler, Jochen: Event-Study-Methodik und statistische Signifikanz, Diss. 2012, Edewecht 2012.

Hollingshead, Andrea, Gwen Wittenbaum, Paul Paulus, Randy Hirokawa, Deborah Ancona, Randall Peterson, Karen Jehn & Kay Yoon: A Look at Groups From the Functional Perspective, in: Poole/Hollingshead, Theories of Small Groups: Interdisciplinary Perspectives, Thousand Oaks 2005, S. 21–62.

Homberg, Fabian & Margit Osterloh: Fusionen und Übernahmen im Licht der Hybris – Überblick über den Forschungsstand, Journal für Betriebswirtschaft 2010 = Bd. 60, S. 269–294.

Hommelhoff, Peter: Eigenkontrolle statt Staatskontrolle – rechtsdogmatischer Überblick zur Aktienrechtsreform 1884, in: Schubert/Hommelhoff, Hundert Jahre modernes Aktienrecht. Eine Sammlung von Texten und Quellen zur Aktienrechtsreform 1884 mit zwei Einführungen, Berlin 1985, S. 53–105.

Hopt, Klaus: Was ist von den Sozialwissenschaften für die Rechtsanwendung zu erwarten?, JuristenZeitung 1975, S. 341–349.

Hopt, Klaus: Comparative Company Law, in: Reimann/Zimmermann, The Oxford Handbook of Comparative Law, Oxford 2008, S. 1161–1191.

Hopt, Klaus & Patrick Leyens: Board Models in Europe – Recent Developments of Internal Governance Structures in Germany, the United Kingdom, France, and Italy, European Company and Financial Law Review 2004, S. 135–168.

Hopt, Klaus & Clemens Trautmann: Zukunftsperspektiven der privatrechtlichen Forschung – Gesellschafts- und Kapitalmarktrecht, Zeitschrift für Unternehmens- und Gesellschaftsrecht 2007 = Bd. 36, S. 480–483.

Horsch, Andreas & Steffen Hundt: Ereignisstudien zu Kapitalmarktreaktionen auf CEO-Turnovers, Corporate Finance biz 2012, S. 225–232.

Howard, George, Scott Maxwell & Kevin Fleming: The proof of the pudding: An illustration of the relative strengths of null hypothesis, meta-analysis, and Bayesian analysis, Psychological Methods 2000 = Bd. 5, S. 315–332.

Howick, Jeremy, Iain Chalmers, Paul Glasziou, Trish Greenhalgh, Carl Heneghan, Alessandro Liberati, Ivan Moschetti, Bob Phillips, Hazel Thornton, Olive Goddard & Mary Hodgkinson: The Oxford 2011 Levels of Evidence (Version 2), Oxford Centre for Evidence-Based Medicine, Zugriff am 16.2.2014 unter www.cebm.net/index.aspx?o=5653.

Hu, Xiaojing: Rechtsfragen der chinesischen Corporate Governance auf Grundlage eines Vergleichs zwischen Deutschland und China, Diss. 2005, Frankfurt/M. 2006.

Hubbard, Raymond & Susie Bayarri: Confusion over Measures of Evidence (p's) versus Errors (α's) in Classical Statistical Testing, The American Statistician 2003 = Bd. 57, S. 171–182.

Huber, Oswald: Das psychologische Experiment. Eine Einführung, 4. Aufl., Bern 2005.

Hübner, Ulrich & Vlad Constantinesco: Einführung in das französische Recht, 3. Aufl., München 1994.

Hüffer, Uwe: Beck'scher Kurzkommentar Aktiengesetz, 10. Aufl., München 2012.

Hume, David: A Treatise of Human Nature: Being an Attempt to Introduce the Experimental Method of Reasoning into Moral Subjects, Hume Texts, Zugriff am 8.11.2012 unter http://www.davidhume.org/texts/thn.html.

Hundt, Steffen & Andreas Horsch: Kapitalmarktreaktionen auf Ankündigungen von M&A-Transaktionen – Eine Ereignisstudie am Beispiel der Unicredit, Corporate Finance biz 2012, S. 141–148.

Hupka, Jan: Das Vergütungsvotum der Hauptversammlung. Eine rechtsökonomische und rechtsvergleichende Untersuchung zu § 120 Abs. 4 AktG, Diss. 2011, Köln 2012.

Husa, Jaako: Classification of Legal Families Today. Is it time for a memorial hymn?, Revue internationale de droit comparé 2004 = Bd. 56, S. 11–38.

Huse, Morten: Accountability and Creating Accountability: a Framework for Exploring Behavioural Perspectives of Corporate Governance, British Journal of Management 2005 = Bd. 16, S. S65–S79.

Hussy, Walter, Margrit Schreier & Gerald Echterhoff: Forschungsmethoden in Psychologie und Sozialwissenschaften für Bachelor, Berlin 2010.

Iacobucci, Frank, Marilyn Pilkington & Robert Prichard: Canadian Business Corporations. An Analysis of Recent Legislative Developments, Agincourt 1977.

Ietswaart, Heleen: Some Notes on the Relations Between Empirical Research and Theory, in: Plett/Ziegert, Empirische Rechtsforschung zwischen Wissenschaft und Politik. Zur Problemlage rechtssoziologischer Auftragsforschung, Tübingen 1984, S. 210–234.

Igarashi, Kiyoshi & Tadanori Makino: Einführung in das japanische Recht, Darmstadt 1990.

Ihrig, Hans-Christoph & Manuel Meder: Zweifelsfragen bei der Zielbenennung zur Zusammensetzung des Aufsichtsrats nach dem Kodex, Zeitschrift für Wirtschaftsrecht 2012, S. 1210–1216.

Ihrig, Hans-Christoph, Andre Wandt & Jonas Wittgens: Die angemessene Vorstandsvergütung drei Jahre nach Inkrafttreten des VorstAG. Grundsätze, Leitlinien und Zweifelsfragen in der Praxis – eine Bestandsaufnahme (Beilage zu Heft 40), Zeitschrift für Wirtschaftsrecht 2012, S. 1–31.

IK Berlin: German Code of Corporate Governance (GCCG), Der Betrieb 2000, S. 1573–1581.

Ilgen, Daniel, John Hollenbeck, Michael Johnson & Dustin Jundt: Teams in Organizations: From Input-Process-Output Models to IMOI Models, Annual Review of Psychology 2005 = Bd. 56, S. 517–543.

Ioannidis, John: Why Most Published Research Findings Are False, Public Library of Science: Medicine 2005 = Bd. 2, S. 696–701.

Isenberg, Daniel: Group Polarization: A Critical Review and Meta-Analysis, Journal of Personality and Social Psychology 1986 = Bd. 50, S. 1141–1151.

Jacquart, Philippe & Scott Armstrong: The Ombudsman: Are Top Executives Paid Enough? An Evidence-Based Review, Interfaces 2013 = Bd. 43, S. 580–589.

Jaeschke, Roman & Gordon Guyatt: What Is Evidence-Based Medicine?, Seminars in Medical Practice 1999 = Bd. 2, S. 3–7.

Jahn, Karoline: Ereignisstudien: Überblick über die Methode, Dipl., Saarbrücken 2007.

Jakob, Raimund & Manfred Rehbinder: Beiträge zur Rechtspsychologie, Berlin 1987.

Janis, Irving: Victims of Groupthink – A Psychological Study of Foreign-policy Decisions and Fiascoes, Boston 1972.

Janka, Bettina: Corporate Governance in Deutschland und Spanien, Diss. 2011, Frankfurt/M. 2011.

Jasso, Guillermina: Factorial Survey Methods for Studying Beliefs and Judgments, Sociological Methods & Research 2006 = Bd. 34, S. 334–423.

Jenkins, Douglas, Atul Mitra, Nina Gupta & Jason Shaw: Are financial incentives related to performance? A meta-analytic review of empirical research, Journal of Applied Psychology 1998 = Bd. 83, S. 777–787.

Jensen, Michael & William Meckling: Theory of the firm: Managerial behavior, agency costs and ownership structure, Journal of Financial Economics 1976 = Bd. 3, S. 305–360.

Jensen, Michael & Kevin Murphy: Performance Pay and Top-Management Incentives, The Journal of Political Economy 1990 = Bd. 98, S. 225–264.

Jiang, Ge: Das GmbH-Recht in China aus rechtsvergleichender Sicht. Analyse, Kritik und Verbesserungsvorschläge, Diss. 2010, Frankfurt/M. 2011.

Jochum, Heike: Grundfragen des Steuerrechts. Eine verfassungsrechtliche und methodische Einführung für Lehre und Praxis, Tübingen 2012.

Jolls, Christine & Cass Sunstein: Debiasing through Law, The Journal of Legal Studies 2006 = Bd. 35, S. 199–242.

Jolls, Christine, Cass Sunstein & Richard Thaler: A Behavioral Approach to Law and Economics, Stanford Law Review 1998 = Bd. 50, S. 1471–1550.

Jones, Dan: A WEIRD View of Human Nature Skews Psychologists' Studies, Science 2010 = Bd. 328, S. 1627.

Jones, Stephen: Was There a Hawthorne Effect?, American Journal of Sociology 1992 = Bd. 98, S. 451–468.

Jost, Fritz: Soziologische Feststellungen in der Rechtsprechung des Bundesgerichtshofs in Zivilsachen, Diss. 1978, Berlin 1979.

Julian, K.: Statisticians—keep it simple, Journal of the Royal Statistical Society. Series D (The Statistician) 1987 = Bd. 36, S. 487–492.

Jung, Heike: Das Kollegialprinzip – ein antizyklisches Thema?, in: Hilgendorf/Rudolf, Festschrift für Wolfgang Heinz, Baden-Baden 2012, S. 883–891.

juris GmbH: Verzeichnis der ausgewerteten juristischen Periodika für die Datenbank Aufsätze, juris Das Rechtsportal > Unternehmen > Auswertung, Zugriff am 3.8.2014 unter http://www.juris.de/jportal/nav/unternehmen/auswertung/fachzeitschriftenaufsaetze.jsp.

Kakabadse, Andrew & Nada Kakabadse (Hg.): Global Boards: One Desire, Many Realities, Basingstoke 2009.

Kalss, Susanne, Christian Nowotny & Martin Schauer (Hg.): Österreichisches Gesellschaftsrecht: systematische Darstellung sämtlicher Rechtsformen. Handbuch, Wien 2008.

Kanazawa, Satoshi: Using laboratory experiments to test theories of corporate behavior, Rationality and Society 1999 = Bd. 11, S. 443–461.

Kanning, Uwe Peter & Anka Hill: Organization-based self-esteem scale – adaptation in an international context, Journal of Business and Media Psychology, Zugriff am 28.8.2012 unter http://journal-bmp.de/?p=1076&lang=en.

Kaplan, Abraham: The Conduct of Inquiry: Methodology for Behavioral Science, San Francisco 1964.

Kaplan, Steven & Joshua Rauh: Wall Street and Main Street: What Contributes to the Rise in the Highest Incomes?, Review of Financial Studies 2010 = Bd. 23, S. 1004–1050.

Karau, Steven & Kipling Williams: Social Loafing: A Meta-Analytic Review and Theoretical Integration, Journal of Personality and Social Psychology 1993 = Bd. 65, S. 681–706.

Karren, Ronald & Melissa Woodard Barringer: A Review and Analysis of the Policy-Capturing Methodology in Organizational Research: Guidelines for Research and Practice, Organizational Research Methods 2002 = Bd. 5, S. 337–361.

Kawamoto, Ichirō: Gesellschaftsrecht in Japan, München 2004.

Kaye, David: Thinking Like a Statistician. The Report of the American Statistical Association Committee on Training in Statistics in Selected Professions, Journal of Legal Education 1984 = Bd. 34, S. 97–102.

Kerr, Norbert: HARKing: Hypothesizing After the Results are Known, Personality and Social Psychology Review 1998 = Bd. 2, S. 196–217.

Kerr, Norbert, Robert MacCoun & Geoffrey Kramer: Bias in Judgment: Comparing Individuals and Groups, Psychological Review 1996 = Bd. 103, S. 687–719.

Kerr, Norbert, Keith Niedermeier & Martin Kaplan: On the Virtues of Assuming Minimal Differences in Information Processing in Individuals and Groups, Group Processes & Intergroup Relations 2000 = Bd. 3, S. 203–217.

Kerr, Norbert & Scott Tindale: Group Performance and Decision Making, Annual Review of Psychology 2004 = Bd. 55, S. 623–655.

KfW: Economic Research: Publikationen, KfW Bankengruppe, Zugriff am 14.10.2012 unter http://www.kfw.de/kfw/de/KfW-Konzern/KfW_Research/Economic_Research/Reihen/index.jsp.

Kiel, Geoffrey & Gavin Nicholson: Board Composition and Corporate Performance: how the Australian experience informs contrasting theories of corporate governance, Corporate Governance: An International Review 2003 = Bd. 11, S. 189–205.

Kim, Peter: When What You Know *Can* Hurt You: A Study of Experiential Effects on Group Discussion and Performance, Organizational Behavior and Human Decision Processes 1997 = Bd. 69, S. 165–177.

Kindler, Peter: Einführung in das italienische Recht: Verfassungsrecht, Privatrecht und internationales Privatrecht, 2. Aufl., München 2008.

Kindler, Peter: Entwicklungslinien des italienischen Gesellschaftsrechts seit Beginn dieses Jahrhunderts, Zeitschrift für Europäisches Privatrecht 2012, S. 72–98.

Kirchner, Hildebert: Bibliographie zum Gesellschaftsrecht. Bücher und Aufsätze aus den Jahren 1965 bis 1970, Zeitschrift für Unternehmens- und Gesellschaftsrecht 1972 = Bd. 1, S. 104–120.

Kirchner, Hildebert: Bibliographie zum Gesellschaftsrecht. Bücher und Aufsätze aus dem Jahre 1991, Zeitschrift für Unternehmens- und Gesellschaftsrecht 1992 = Bd. 21, S. 648–716.

Kirchner, Hildebert & Dietrich Pannier: Abkürzungsverzeichnis der Rechtssprache, 7. Aufl., Berlin 2012.

Kissel, Otto Rudolf: Der dreistufige Aufbau in der ordentlichen Gerichtsbarkeit. Ein Beitrag zur Großen Justizreform, Frankfurt/M. 1972.

Klausing, Friedrich: Aktien-Gesetz nebst Einführungsgesetz und „Amtlicher Begründung“, Berlin 1937.

Klein, April: Firm Performance and Board Committee Structure, Journal of Law and Economics 1998 = Bd. 41, S. 275–304.

Klick, Jonathan: The Perils of Empirical Work on Institutions, Journal of Institutional and Theoretical Economics 2010 = Bd. 166, S. 166–170.

Klimoski, Richard & Susan Mohammed: Team Mental Model: Construct or Metaphor?, Journal of Management 1994 = Bd. 20, S. 403–437.

Klöhn, Lars: Kapitalmarkt, Spekulation und Behavioral Finance. Eine interdisziplinäre und vergleichende Analyse zum Fluch und Segen der Spekulation und ihrer Regulierung durch Recht und Markt, Diss. 2005, Berlin 2006.

Kluth, Winfried & Ferdinand Goltz: Führungsteam statt Hauptgeschäftsführer, Gewerbearchiv 2003, S. 265–271.

Köbler, Gerhard: Deutsches Etymologisches Wörterbuch 1995, Zentrissimum integrativer europäischer Legistik (Innsbruck), Zugriff am 3.8.2014 unter www.koeblergerhard.de/derwbhin.html.

Köbler, Gerhard: Lateinisches Abkunfts- und Wirkungswörterbuch (2. Aufl.), Zentrissimum integrativer europäischer Legistik (Innsbruck), Zugriff am 29.6.2012 unter http://www.koeblergerhard.de/Latein2/LAWVorwort2.html.

Koch, Horst Heinrich & Hans-Peter Steinmetz: Gegen Einzelrichter im Verwaltungsprozeß, Die Öffentliche Verwaltung 1981, S. 50–55.

Kocher, Dirk: Kapitalgesellschaft & Co. KG in Finnland. Eine Untersuchung zur Frage der Übertragbarkeit der deutschen Grundtypenvermischung auf das finnische Recht, Diss. 2003, Frankfurt/M. 2003.

Koenige, Heinrich, Robert Teichmann & Walter Koehler: Handausgabe des Handelsgesetzbuchs vom 10. Mai 1897, 4. Aufl., Berlin 1936.

Koh Soon Kwang, Peter: Major Issues in Company Law, Singapore 2010.

Kohl, Christian, Marc Steffen Rapp & Michael Wolff: Kodexakzeptanz 2011: Analyse der Entsprechenserklärungen zum DCGK, Der Aufsichtsrat 2011, S. 108–110.

Kopp, Ferdinand: Welchen Anforderungen soll eine einheitliche Verwaltungsprozeßordnung genügen, um im Rahmen einer funktionsfähigen Rechtspflege effektiven Rechtsschutz zu gewährleisten?, in: DJT, Gutachten für den 54. Deutschen Juristentag, München 1982, S. B1 ff.

Koriat, Asher: When Are Two Heads Better than One and Why?, Science 2012 = Bd. 336, S. 360–362.

Köritz, Anja: Konvergenz und Divergenz der Corporate Governance in Deutschland und Großbritannien: ein Rechtsvergleich, Diss. 2010, Baden-Baden 2010.

Kornblum, Udo: Rechtstatsachen zum Unternehmens- und Gesellschaftsrecht, GmbH-Rundschau 1981, S. 227–235.

Kornblum, Udo: Bundesweite Rechtstatsachen zum Unternehmens- und Gesellschaftsrecht (Stand 1.1.2012), GmbH-Rundschau 2012, S. 728–735.

Kornblum, Udo: Bundesweite Rechtstatsachen zum Unternehmens- und Gesellschaftsrecht (Stand 1.1.2013), GmbH-Rundschau 2013, S. 693–703.

Körner, Alois: Gedanken zum Kollegialprinzip, Zeitschrift für Öffentliches Recht 1937 = Bd. 17, S. 55–63.

Kort, Michael: Das „Mannesmann"-Urteil im Lichte von § 87 AktG, Neue Juristische Wochenschrift 2005, S. 333–336.

Kothari, S. P. & Jerold Warner: Chapter 1. Econometrics of Event Studies, in: Eckbo, Handbook of Corporate Finance. Empirical Corporate Finance, San Diego 2007, S. 3–36.

Kötz, Hein & Hans-Bernd Schäfer: Schadensverhütung durch ökonomische Anreize: Eine empirische Untersuchung, Archiv für die civilistische Praxis 1989 = Bd. 189, S. 501–525.

Kovner, Anthony, Jeffrey Elton & John Billings: Evidence-based management, Frontiers of Health Services Management 2000 = Bd. 16, S. 3–24.

Kozlowski, Steve & Daniel Ilgen: Enhancing the Effectiveness of Work Groups and Teams, Psychological Science in the Public Interest 2006 = Bd. 7, S. 77–124.

Kraakman, Reinier, John Armour, Paul Davies, Luca Enriques, Henry Hansmann, Gerard Hertig, Klaus Hopt, Hideki Kanda & Edward Rock: The Anatomy of Corporate Law. A Comparative and Functional Approach, 2. Aufl., Oxford 2009.

Krämer, Walter: Statistik: Vom Geburtshelfer zum Bremser der Erkenntnis in den Sozialwissenschaften?, Kölner Zeitschrift für Soziologie und Sozialpsychologie. Sonderheft 44: Methoden der Sozialforschung 2004, S. 51–60.

Krämer, Walter: Statistik verstehen. Eine Gebrauchsanweisung, 9. Aufl., München 2010.

Krämer, Walter: Denkste! Trugschlüsse aus der Welt des Zufalls und der Zahlen, Neuausg., München 2011.

Krämer, Walter: So lügt man mit Statistik, Neuausg., München 2011.

Krauss, Lawrence: „Nehmen wir an, die Kuh ist eine Kugel…“ Nur keine Angst vor Physik, München 1998.

Kravitz, David & Barbara Martin: Ringelmann rediscovered: The original article, Journal of Personality and Social Psychology 1986 = Bd. 50, S. 936–941.

Krejci, Heinz: Vereinsgesetz 2002. Manzscher Kurzkommentar, Wien 2002.

Kritzer, Herbert: The (Nearly) Forgotten Early Empirical Legal Research, in: Cane/Kritzer, The Oxford Handbook of Empirical Legal Research, Oxford 2010, S. 875–900.

Krohn, Roger: Why are graphs so central in science?, Biology and Philosophy 1991 = Bd. 6, S. 181–203.

Kronmal, Richard: Spurious Correlation and the Fallacy of the Ratio Standard Revisited, Journal of the Royal Statistical Society. Series A (Statistics in Society) 1993 = Bd. 156, S. 379–392.

Kropff, Bruno: Aktiengesetz (Textausgabe und Materialien), Düsseldorf 1965.

Kroymann, Benjamin: Das Kapitalgesellschaftsrecht der VR China. Analyse der Rahmenbedingungen für ausländische Investoren, Diss. 2008, Tübingen 2009.

Krugmann, Michael: Evidenzfunktionen, Diss. 1995, Berlin 1996.

Kugler, Tamar, Edgar Kausel & Martin Kocher: Are groups more rational than individuals? A review of interactive decision making in groups, Wiley Interdisciplinary Reviews: Cognitive Science 2012 = Bd. 3, S. 471–482.

Kühberger, Anton: The Influence of Framing on Risky Decisions: A Meta-analysis, Organizational Behavior and Human Decision Processes 1998 = Bd. 75, S. 23–55.

Kunz, Jennifer & Stefan Linder: ZP-Stichwort: Vignetten-Experiment, Zeitschrift für Planung & Unternehmenssteuerung 2011 = Bd. 21, S. 211–222.

Kuper, Magdalena: Die Aktiengesellschaft in Polen und Deutschland: rechtsvergleichende Analyse der Organisationsstrukturen im Lichte der europäischen Rechtsangleichung, Diss., 2005.

Kuselit Verlag: Gesamtliste der von Kuselit dokumentierten Zeitschriften, Kuselit Online, Zugriff am 3.8.2014 unter www.kuselit.de/cms/index.php?folder=1041&op=gesamtliste.

Lamnek, Siegfried: Qualitative Sozialforschung: Lehrbuch, 5. Aufl., Weinheim 2010.

Langenbucher, Katja: Zentrale Akteure der Corporate Governance: Zusammensetzung des Aufsichtsrats. Zum Vorschlag einer obligatorischen Besetzungserklärung, Zeitschrift für Unternehmens- und Gesellschaftsrecht 2012, S. 314–342.

Langevoort, Donald: Organized Illusions: A Behavioral Theory of Why Corporations Mislead Stock Market Investors (And Cause Other Social Harms), University of Pennsylvania Law Review 1997 = Bd. 146, S. 101–172.

Langevoort, Donald: The Human Nature of Corporate Boards: Law, Norms, and the Unintended Consequences of Independence and Accountability, The Georgetown Law Journal 2001 = Bd. 89, S. 797–832.

Langevoort, Donald: Heuristics Inside the Firm – Perspectives from Behavioral Law and Economics, in: Gigerenzer/Engel, Heuristics and the Law, Boston 2006, S. 87–100.

Larenz, Karl & Claus-Wilhelm Canaris: Methodenlehre der Rechtswissenschaft, 3. Aufl. 1995, Nachdruck, Berlin 2008.

Larson, James: In Search of Synergy in Small Group Performance, New York 2010.

Laughlin, Patrick: Social combination processes of cooperative problem-solving groups on verbal intellective tasks, in: Fishbein, Progress in Social Psychology, Hillsdale 1980, S. 127–155.

Laughlin, Patrick: Group Problem Solving, Princeton 2011.

Lavine, Samuel: The Business Corporations Act. An Analysis, Toronto 1971.

Lawless, Robert, Jennifer Robbennolt & Thomas Ulen: Empirical Methods in Law, Frederick 2010.

Lazarsfeld, Paul: The American Soldier—An Expository Review, Public Opinion Quarterly 1949 = Bd. 13, S. 377–404.

Lazear, Edward: Economic Imperialism, The Quarterly Journal of Economics 2000 = Bd. 115, S. 99–146.

Leary, Mark: Introduction to behavioral research methods, 5. Aufl., Boston 2008.

Lee, Hyeong-Kyu: Die Geschäftsführung und Kontrolle in der Aktiengesellschaft. Eine rechtsvergleichende Untersuchung nach koreanischem und deutschem Aktienrecht, Diss. 1990, Göttingen 1991.

Lee, Ki-Su: Corporate Governance im koreanischen Gesellschaftsrecht, in: Dauner-Lieb/Hommelhoff/Jacobs/Kaiser/Weber, Festschrift für Horst Konzen zum siebzigsten Geburtstag, Tübingen 2006, S. 491–500.

Legewie, Joscha: Die Schätzung von kausalen Effekten: Überlegungen zu Methoden der Kausalanalyse anhand von Kontexteffekten in der Schule, Kölner Zeitschrift für Soziologie und Sozialpsychologie 2012 = Bd. 64, S. 123–153.

Lehrer, Jonah: The Truth Wears Off, The New Yorker, Zugriff am 21.6.2012 unter http://www.newyorker.com/reporting/2010/12/13/101213fa_fact_lehrer.

Lempert, Richard: The Significance of Statistical Significance, Law & Social Inquiry 2009 = Bd. 34, S. 225–249.

Lempert, Richard: Empirical Research for Public Policy: With Examples from Family Law, Journal of Empirical Legal Studies 2010 = Bd. 5, S. 907–926.

Lenhard, Johannes: Models and Statistical Inference: The Controversy between Fisher and Neyman–Pearson, The British Journal for the Philosophy of Science 2006 = Bd. 57, S. 69–91.

Lenoble, Jacques: From an Incentive to a Reflexive Approach to Corporate Governance, in: Cobbaut/Lenoble, Corporate Governance: An Institutionalist Approach, Den Haag 2003, S. 17–63.

Lercher, Josef: Methoden der Rechtstatsachenforschung – Ein Abriss, in: Barta/Ganner/Lichtmannegger, Rechtstatsachenforschung – Heute. Tagungsband 2008, Innsbruck 2009, S. 205–240.

Lerdal, Susan Nissen: Evidence-Based Librarianship: Opportunity for Law Librarians?, Law Library Journal 2006 = Bd. 98, S. 33–60.

Leu, Hans & Ekkehard Werner: Der Einzelne und die Gruppe im juristischen Entscheidungsprozess, in: Bender, Tatsachenforschung in der Justiz. Einführung in die Methoden, Tübingen 1972, S. 127–142.

Levine, John & Richard Moreland: Progress in Small Group Research, Annual Review of Psychology 1990 = Bd. 41, S. 585–634.

Levine, Kay: The Law is not the Case: Incorporating Empirical Methods into the Culture of Case Analysis, University of Florida Journal of Law and Public Policy 2006 = Bd. 17, S. 283–301.

Levinson, Harry: Why the Behemoths Fell. Psychological Roots of Corporate Failure, American Psychologist 1994 = Bd. 49, S. 428–436.

Levitt, Steven & John List: Was There Really a Hawthorne Effect at the Hawthorne Plant? An Analysis of the Original Illumination Experiments, American Economic Journal: Applied Economics 2011 = Bd. 3, S. 224–238.

LG Düsseldorf: Urt. v. 14.7.2010, Az. 14 KLs 6/09, Die Aktiengesellschaft 2011, S. 722–723.

Liebs, Detlef: Lateinische Rechtsregeln und Rechtssprichwörter, 7. Aufl., München 2007.

Lieder, Jan: Über den Nutzen der Rechtstatsachenforschung zur Verbesserung der Corporate Governance unter besonderer Berücksichtigung historischer Studien, in: Bayer, Die Aktiengesellschaft im Spiegel der Rechtstatsachenforschung, Jena 2007, S. 79–110.

Lighthall, William Schuyler: The Dominion Companies Act 1934 annotated, Montreal 1935.

Limbach, Jutta: Theorie und Wirklichkeit der GmbH. Die empirischen Normaltypen der GmbH und ihr Verhältnis zum Postulat von Herrschaft und Haftung, Berlin 1966.

Limbach, Jutta: Empirische Rechtssoziologie. Aufgabe, Untersuchungsgegenstand, Methode, Juristische Arbeitsblätter 1973, S. 143–151 und 295–300.

Lin, Yueh-hsiang, Shing-yang Hu & Ming-shen Chen: Managerial optimism and corporate investment: Some empirical evidence from Taiwan, Pacific-Basin Finance Journal 2005 = Bd. 13, S. 523–546.

Lipsey, Mark & David Wilson: The Efficacy of Psychological, Educational, and Behavioral Treatment. Confirmation From Meta-Analysis, American Psychologist 1993 = Bd. 48, S. 1181–1209.

List, John: *Homo experimentalis* Evolves, Science 2008 = Bd. 321, S. 207–208.

Listokin, Yair: Management Always Wins the Close Ones, American Law and Economics Review 2008 = Bd. 10, S. 159–184.

Loewenstein, George: Experimental Economics From the Vantage-Point of Behavioural Economics, The Economic Journal 1999 = Bd. 109, S. 25–34.

Loftus, Geoffrey: A picture is worth a thousand p values: On the irrelevance of hypothesis testing in the microcomputer age, Behavior Research Methods 1993 = Bd. 25, S. 250–256.

Lorenz, Jörn, Manja Pietzcker & Frank Pietzcker: Empirische Sprachgebrauchsanalyse – Entlarvt ein neues Beweismittel Verletzungen des Analogieverbots (Art. 103 II GG)?, Neue Zeitschrift für Strafrecht 2005, S. 429–434.

Lorge, Irving, David Fox, Joel Davitz & Marlin Brenner: A Survey of Studies Contrasting the Quality of Group Performance and Individual Performance, 1920–1957, Psychological Bulletin 1958 = Bd. 55, S. 337–372.

Lorse, Jürgen: Kollegiale Führungsstrukturen in öffentlichen Verwaltungen – Rechtsgrundlagen, Gestaltungsperspektiven in der Verwaltungspraxis, Zeitschrift für Beamtenrecht 2003, S. 185–198.

Lu, Li, Connie Yuan & Poppy Lauretta McLeod: Twenty-Five Years of Hidden Profiles in Group Decision Making, Personality and Social Psychology Review 2012 = Bd. 16, S. 54–75.

Lüdemann, Jörn: Die Grenzen des homo oeconomicus und die Rechtswissenschaft, in: Engel/Englerth/Lüdemann/Spiecker gen. Döhmann, Recht und Verhalten. Beiträge zu Behavioral Law and Economics, Tübingen 2007, S. 7–59.

Lüdemann, Jörn: Jenseits von Rationalität und Eigennutz, in: Engel/Englerth/Lüdemann/ Spiecker gen. Döhmann, Recht und Verhalten. Beiträge zu Behavioral Law and Economics, Tübingen 2007, S. 1–4.

Lüdemann, Volker: Das Recht der Aktiengesellschaft in Rußland: von den Anfängen bis zum Aktiengesetz von 1996, Diss. 2000, Osnabrück 2001.

Lüderssen, Klaus: Erfahrung als Rechtsquelle: Abduktion und Falsifikation von Hypothesen im juristischen Entscheidungsprozess. Eine Fallstudie aus dem Kartellrecht, Habil. 1970, Frankfurt/M. 1972.

Lutter, Marcus: Bankenvertreter im Aufsichtsrat, Zeitschrift für das gesamte Handels- und Wirtschaftsrecht 1981 = Bd. 145, S. 224–251.

Lutter, Marcus: Professionalisierung der Aufsichtsräte, Neue Juristische Wochenschrift 1995, S. 1133–1134.

Lutter, Marcus: Der Letter of Intent. Zur rechtlichen Bedeutung von Absichtserklärungen, 3. Aufl., Köln 1998.

Lutter, Marcus: Corporate Governance und ihre aktuellen Probleme, vor allem: Vorstandsvergütung und ihre Schranken, Zeitschrift für Wirtschaftsrecht 2003, S. 737–743.

Lutter, Marcus: Professionalisierung des Aufsichtsrats, Der Betrieb 2009, S. 775–779.

Lutter, Marcus & Peter Hommelhoff (Hg.): SE-Kommentar: SE-VO, SEAG, SEBG, Steuerrecht, Köln 2008.

Lutter, Marcus, Alexander Kollmorgen & Heiner Feldhaus: Die Europäische Aktiengesellschaft – Satzungsgestaltung bei der „mittelständischen SE“, Betriebs-Berater 2005, S. 2473–2483.

Lutter, Marcus & Gerd Krieger: Rechte und Pflichten des Aufsichtsrats, 5. Aufl., Köln 2008.

Lutter, Marcus & Karsten Schmidt (Hg.): Aktiengesetz, 2 Bde., 2. Aufl., Köln 2010.

MacCoun, Robert & Norbert Kerr: Asymmetric influence in mock jury deliberation: Jurors' bias for leniency, Journal of Personality and Social Psychology 1988 = Bd. 54, S. 21–33.

MacKenzie, Doris Layton: Evidence-Based Corrections: Identifying What Works, Crime & Delinquency 2000 = Bd. 46, S. 457–471.

MacKinlay, Craig: Event Studies in Economics and Finance, Journal of Economic Literature 1997 = Bd. 35, S. 13–39.

Mahoney, Michael: Publication prejudices: An experimental study of confirmatory bias in the peer review system, Cognitive Therapy and Research 1977 = Bd. 1, S. 161–175.

Mäki, Uskali: Economics Imperialism: Concept and Constraints, Philosophy of the Social Sciences 2009 = Bd. 39, S. 351–380.

Malmendier, Ulrike & Geoffrey Tate: CEO Overconfidence and Corporate Investment, The Journal of Finance 2005 = Bd. 60, S. 2661–2700.

Malmendier, Ulrike & Geoffrey Tate: Does Overconfidence Affect Corporate Investment? CEO Overconfidence Measures Revisited, European Financial Management 2005 = Bd. 11, S. 649–659.

Malmendier, Ulrike & Geoffrey Tate: Who makes acquisitions? CEO overconfidence and the market's reaction, Journal of Financial Economics 2008 = Bd. 89, S. 20–43.

Mann, Charles: Can meta-analysis make policy?, Science 1994 = Bd. 266, S. 960–962.

Markman, Arthur, Todd Maddox & Darrell Worthy: Choking and Excelling Under Pressure, Psychological Science 2006 = Bd. 17, S. 944–948.

Martens, Klaus-Peter: Der Grundsatz gemeinsamer Vorstandsverantwortung, in: Goerdeler/Hommelhoff/Lutter/Wiedemann, Festschrift für Hans-Joachim Fleck zum 70. Geburtstag am 30. Januar 1988, Berlin 1988, S. 191–208.

Martin, Jens: The Impact of Optimistic and Privately Informed Managers on Firm Performance and Corporate Decisions, Università della Svizzera italiana, Zugriff am 15.3.2012 unter http://www.common.unisi.ch/pdf_pub3894_2.

Marutschke, Hans Peter: Einführung in das japanische Recht, 2. Aufl., München 2010.

Maslow, Abraham: The Psychology of Science. A Reconnaissance, New York 1966.

Masten, Cornelius Arthur & William Kaspar Fraser: Company Law of Canada, 3. Aufl., Toronto 1929.

Mastronardi, Philippe: Juristisches Denken. Eine Einführung, 2. Aufl., Bern 2003.

Mathieu, John, Travis Maynard, Tammy Rapp & Lucy Gilson: Team Effectiveness 1997-2007: A Review of Recent Advancements and a Glimpse Into the Future, Journal of Management 2008 = Bd. 34, S. 410–476.

Mathieu, Sylvain, Isabelle Boutron, David Moher, Douglas Altman & Philippe Ravaud: Comparison of Registered and Published Primary Outcomes in Randomized Controlled Trials, JAMA: The Journal of the American Medical Association 2009 = Bd. 302, S. 977–984.

Matthäus, Wolf-Gert: Statistische Tests mit Excel leicht erklärt. Beurteilende Statistik für jedermann, Wiesbaden 2007.

Matthews, Robert: Storks Deliver Babies (p= 0.008), Teaching Statistics 2000 = Bd. 22, S. 36–38.

May, Axel: Zum Stand der empirischen Forschung über Informationsverarbeitung am Aktienmarkt – Ein Überblick, Zeitschrift für betriebswirtschaftliche Forschung 1991 = Bd. 43, S. 313–335.

Mayer-Maly, Theo: Der Jurist und die Evidenz, in: Marcic/Mosler/Suy/Zemanek, Internationale Festschrift für Alfred Verdroß zum 80. Geburtstag, München 1971, S. 259–270.

Mayring, Philipp: Einführung in die qualitative Sozialforschung: Eine Anleitung zu qualitativem Denken, 5. Aufl., Weinheim 2002.

McCloskey, Deirdre: The Rhetoric of Economics, 2. Aufl., Madison 1998.

McCloskey, Donald: The Loss Function Has Been Mislaid: The Rhetoric of Significance Tests, American Economic Review 1985 = Bd. 75, S. 201–205.

McDermott, Rose: Experimental Methods in Political Science, Annual Review of Political Science 2002 = Bd. 5, S. 31–61.

McGrath, Joseph & David Kravitz: Group Research, Annual Review of Psychology 1982 = Bd. 33, S. 195–230.

McKelvey, Bill & Pierpaolo Andriani: Why Gaussian statistics are mostly wrong for strategic organization, Strategic Organization 2005 = Bd. 3, S. 219–228.

McWilliams, Abagail & Donald Siegel: Event Studies in Management Research: Theoretical and Empirical Issues, The Academy of Management Journal 1997 = Bd. 40, S. 626–657.

Meckel, Verena: Die Corporate Governance im neuen japanischen Gesellschaftsrecht unter besonderer Berücksichtigung der Aufgaben von Verwaltungs- und Prüferrat, Diss. 2009, Tübingen 2010.

Menjucq, Michel: Das „monistische“ System der Unternehmensleitung in der SE, Zeitschrift für Unternehmens- und Gesellschaftsrecht 2003, S. 679–687.

Mense, Christian & Erik Rosenhäger: Mehr Vielfalt wagen – Zu den jüngsten Änderungen des Deutschen Corporate Governance Kodex, Gesellschafts- und Wirtschaftsrecht 2010, S. 311–314.

Merchant, Kenneth, Wim Van der Stede, Thomas Lin & Zengbiao Yu: Performance Measurement and Incentive Compensation: An Empirical Analysis and Comparison of Chinese and Western Firms’ Practices, European Accounting Review 2011 = Bd. 20, S. 639–667.

Merkt, Hanno: Die monistische Unternehmensverfassung für die Europäische Aktiengesellschaft aus deutscher Sicht. Mit vergleichendem Blick auf die Schweiz, das Vereinigte Königreich und Frankreich, Zeitschrift für Unternehmens- und Gesellschaftsrecht 2003, S. 650–678.

Merkt, Hanno: Die Zukunft der privatrechtlichen Forschung im Unternehmens- und Kapitalmarktrecht, Zeitschrift für Unternehmens- und Gesellschaftsrecht 2007 = Bd. 36, S. 532–541.

Merkt, Hanno & Stephan Göthel: US-amerikanisches Gesellschaftsrecht, 2. Aufl., Frankfurt/M. 2006.

Merz, Christian & Andreas Wömpener: Zur Relevanz des Enforcements in Deutschland – Der Fall der fehlenden Prognose der Merck KGaA, Corporate Finance biz 2012, S. 268–271.

Mesmer-Magnus, Jessica & Leslie DeChurch: Information sharing and team performance: A meta-analysis, Journal of Applied Psychology 2009 = Bd. 94, S. 535–546.

Meyer, Justus: Die GmbH und andere Handelsgesellschaften im Spiegel empirischer Forschung (I), GmbH-Rundschau 2002, S. 177–189.

Meyer, Justus: Die GmbH und andere Handelsgesellschaften im Spiegel empirischer Forschung (II), GmbH-Rundschau 2002, S. 242–256.

Mezulis, Amy, Lyn Abramson, Janet Hyde & Benjamin Hankin: Is There a Universal Positivity Bias in Attributions? A Meta-Analytic Review of Individual, Developmental, and Cultural Differences in the Self-Serving Attributional Bias, Psychological Bulletin 2004 = Bd. 130, S. 711–747.

Michaelsen, Larry, Warren Watson & Robert Black: A realistic test of individual versus group consensus decision making, Journal of Applied Psychology 1989 = Bd. 74, S. 834–839.

Michaelsen, Larry, Warren Watson, Albert Schwartzkopf & Robert Black: Group decision making: How you frame the question determines what you find, Journal of Applied Psychology 1992 = Bd. 77, S. 106–108.

Michalski, Lutz (Hg.): Kommentar zum Gesetz betreffend die Gesellschaften mit beschränkter Haftung (GmbH-Gesetz), 2 Bde., 2. Aufl., München 2010.

Miettinen, Johanna: Die Reform des finnischen Aktienrechts, Recht der internationalen Wirtschaft 2006, S. 812–815.

Mintken, Karl-Heinz: Empirische Untersuchungen als Seminararbeiten, Verwaltungsrundschau 1992, S. 252–253.

Mitchell, Gregory: Libertarian Paternalism is an Oxymoron, Northwestern University Law Review 2005 = Bd. 99, S. 1245–1278.

Mitchell, Gregory: The Promise and Limitations of an Empirical Approach to Law, Virginia Journal 2008 = Bd. 11, S. 28–46.

Mitchell, Gregory: Revisiting Truth or Triviality: The External Validity of Research in the Psychological Laboratory, Perspectives on Psychological Science 2012 = Bd. 7, S. 109–117.

Mohammed, Susan, Lori Ferzandi & Katherine Hamilton: Metaphor No More: A 15-Year Review of the Team Mental Model Construct, Journal of Management 2010 = Bd. 36, S. 876–910.

MOLEG: Commercial Act (Republic of Korea), Ministry of Government Legislation: Korean Laws in English, Zugriff am 8.10.2012 unter http://www.moleg.go.kr/FileDownload.mo?flSeq=31318.

Monahan, John & Laurens Walker: Empirical Questions Without Empirical Answers, Wisconsin Law Review 1991, S. 569–594.

Mook, Douglas: In defense of external invalidity, American Psychologist 1983 = Bd. 38, S. 379–387.

Moye, John: The law of business organizations, 6. Aufl., Clifton Park 2004.

Mullen, Brian & Catherine Riordan: Self-Serving Attributions for Performance in Naturalistic Settings: A Meta-Analytic Review1, Journal of Applied Social Psychology 1988 = Bd. 18, S. 3–22.

Müller, Friedrich & Ralph Christensen: Juristische Methodik Bd. 1: Grundlegung für die Arbeitsmethoden der Rechtspraxis, 11. Aufl., Berlin 2013.

Münzel, Frank: 27.10.05/1 Gesellschaftsgesetz, Chinas Recht. Kommentierte Übersetzungen aus dem Recht der Volksrepublik China, Zugriff am 4.10.2012 unter http://lehrstuhl.jura.uni-goettingen.de/chinarecht/2005.zip.

Murnighan, Keith & Alvin Roth: Effects of group size and communication availability on coalition bargaining in a veto game, Journal of Personality and Social Psychology 1980 = Bd. 39, S. 92–103.

Narby, Douglas, Brian Cutler & Gary Moran: A meta-analysis of the association between authoritarianism and jurors' perceptions of defendant culpability, Journal of Applied Psychology 1993 = Bd. 78, S. 34–42.

Nard, Craig Allen: Empirical Legal Scholarship: Reestablishing A Dialogue Between The Academy And Profession, Wake Forest Law Review 1995 = Bd. 30, S. 347–368.

Neck, Christopher: Letterman or Leno: a groupthink analysis of successive decisions made by the National Broadcasting Company (NBC), Journal of Managerial Psychology 1996 = Bd. 11, S. 3–17.

Nemeth, Charlan, Keith Brown & John Rogers: Devil's advocate versus authentic dissent: stimulating quantity and quality, European Journal of Social Psychology 2001 = Bd. 31, S. 707–720.

Neumann, Richard & Stefan Krieger: Empirical Inquiry Twenty-Five Years After *The Lawyering Process*, Clinical Law Review 2003 = Bd. 10, S. 349–397.

Nickerson, Raymond: Null Hypothesis Significance Testing: A Review of an Old and Continuing Controversy, Psychological Methods 2000 = Bd. 5, S. 241–301.

Nicolaysen, Isaschar: Das neue schwedische Aktiengesetz, Recht der internationalen Wirtschaft 2005, S. 884–894.

Nielsen, Laura Beth: The Need for Multi-Method Approaches in Empirical Legal Research, in: Cane/Kritzer, The Oxford Handbook of Empirical Legal Research, Oxford 2010, S. 951–975.

Niemeier, Wilhelm: Rechtstatsachen und Rechtsfragen der Einziehung von GmbH-Anteilen, Diss. 1981, Heidelberg 1982.

Niemeier, Wilhelm: GmbH und Limited im Markt der Unternehmensrechtsträger. Marktdaten zur Reformdebatte, Zeitschrift für Wirtschaftsrecht 2006, S. 2237–2250.

Niemeier, Wilhelm: Die „Mini-GmbH" (UG) trotz Marktwende bei der Limited?, Zeitschrift für Wirtschaftsrecht 2007, S. 1794–1801.

Niemeier, Wilhelm: „Triumph" und Nachhaltigkeit deutscher Ein-Euro-Gründungen – Rechtstatsachen zur Limited und ein Zwischenbericht zur Unternehmensgesellschaft, in: Altmeppen/Fitz/Honsell, Festschrift für Günter H. Roth zum 70. Geburtstag, München 2011, S. 533–551.

Nowak, Eric, Roland Rott & Till Mahr: Wer den Kodex nicht einhält, den bestraft der Kapitalmarkt? Eine empirische Analyse der Selbstregulierung und Kapitalmarktrelevanz des Deutschen Corporate Governance Kodex, Zeitschrift für Unternehmens- und Gesellschaftsrecht 2005, S. 252–279.

Nußbaum, Arthur: Die Rechtstatsachenforschung, Archiv für die civilistische Praxis 1955 = Bd. 154, S. 453–484.

Nußbaum, Arthur: Die Rechtstatsachenforschung. Programmschriften und praktische Beispiele, Berlin 1968.

Nussbaum, Arthur: Die Rechtstatsachenforschung. Ihre Bedeutung für Wissenschaft und Unterricht, Tübingen 1914.

Nussbaum, Arthur: Fact Research in Law, Columbia Law Review 1940 = Bd. 40, S. 189–219.

NYSE: Listed Company Manual, New York Stock Exchange, Zugriff am 11.10.2012 unter http://nysemanual.nyse.com/lcm/.

o.V.: Evidence-based morality, Canadian Medical Association Journal 1999 = Bd. 161, S. 229.

O'Connor, Marleen: The Enron Board: The Perils of Groupthink, University of Cincinnati Law Review 2003 = Bd. 71, S. 1233–1320.

Oesterle, Michael-Jörg: Entscheidungsfindung im Vorstand großer deutscher Aktiengesellschaften, Zeitschrift Führung + Organisation 2003 = Bd. 72, S. 199–208

Oestmann, Peter: Die Ermittlung von Verkehrssitten und Handelsbräuchen im Zivilprozeß, JuristenZeitung 2003, S. 285–290.

Ooi, Yao Hua: Simpson's Paradox – A Survey of Past, Present and Future Research, Wharton Research Scholars Journal (U. Penn. ScholarlyCommons) 2004, Zugriff am 27.6.2014 unter http://repository.upenn.edu/wharton_research_scholars/15.

Open Science Collaboration: An Open, Large-Scale, Collaborative Effort to Estimate the Reproducibility of Psychological Science, Perspectives on Psychological Science 2012 = Bd. 7, S. 657–660.

Oplustil, Krzysztof: § 15. Kapitalgesellschaftsrecht, in: Liebscher/Zoll, Einführung in das polnische Recht, München 2005, S. 410–444.

Otto, Hans-Uwe, Andreas Polutta & Holger Ziegler: What Works? Welches Wissen braucht die Soziale Arbeit? Zum Konzept evidenzbasierter Praxis, Opladen 2010.

Palfrey, Thomas: Laboratory Experiments, in: Wittman/Weingast, The Oxford Handbook of Political Economy, Oxford 2006, S. 915–936.

Palfrey, Thomas: Laboratory Experiments in Political Economy, Annual Review of Political Science 2009 = Bd. 12, S. 379–388.

Palmer, Earl: Directors' Powers and Duties, in: Ziegel, Studies in Canadian Company Law (Etudes sur le droit Canadien des compagnies) vol. 1, Toronto 1967, S. 365.

Palmer, Francis Beaufort & Clive Maximilian Schmitthoff: Palmer's Company Law, 24. Aufl., London 1987.

Paredes, Troy: Too Much Pay, Too Much Deference: Behavioral Corporate Finance, CEOs, and Corporate Governance, Florida State University Law Review 2005 = Bd. 32, S. 673–762.

Parlament, Republik Österreich: Nationalrat, XXI. Gesetzgebungsperiode, Regierungsvorlage 990 d.B. („Bundesgesetz über Vereine"), Materialien, Parlament aktiv: Regierungsvorlagen und Gesetzesinitiativen, Zugriff am 5.10.2012 unter http://www.parlament.gv.at/PAKT/VHG/XXI/I/I_00990/fname_000243.pdf.

Parsons, Mcllvaine: What Happened at Hawthorne?, Science 1974 = Bd. 183, S. 922–932.

Pashchenko, Tatiana: § 21. Gesellschaftsrecht, in: Nußberger, Einführung in das russische Recht, München 2010, S. 216–235.

Pavlides, Marios & Michael Perlman: How Likely Is Simpson's Paradox?, The American Statistician 2009 = Bd. 63, S. 226–233.

Payne, Tyge, George Benson & David Finegold: Corporate Board Attributes, Team Effectiveness and Financial Performance, Journal of Management Studies 2009 = Bd. 46, S. 704–731.

Peifer, Markus: Das Recht der Kapitalgesellschaften in Frankreich, Berlin 2009.

Peltokorpi, Vesa: Transactive memory systems, Review of General Psychology 2008 = Bd. 12, S. 378–394.

Peltzer, Martin: Das Grünbuch der EU-Kommission vom 5. 4. 2011 und die Deutsche Corporate Governance, Neue Zeitschrift für Gesellschaftsrecht 2011, S. 961–968.

Pepper, Alexander, Julie Gore & Alf Crossman: Are long-term incentive plans an effective and efficient way of motivating senior executives?, Human Resource Management Journal 2013 = Bd. 23, S. 36–51.

Peters, Douglas & Stephen Ceci: Peer-review practices of psychological journals: The fate of published articles, submitted again, Behavioral and Brain Sciences 1982 = Bd. 5, S. 187–195.

Petersen, Niels: Braucht die Rechtswissenschaft eine empirische Wende?, Der Staat 2010, S. 435–455.

Petersen, Niels & Sebastian Goerg: § 8. Empirische Methoden, in: Towfigh/Petersen, Ökonomische Methoden im Recht, Tübingen 2010, S. 201–241.

Peterson, Pamela: Event Studies: A Review of Issues and Methodology, Quarterly Journal of Business and Economics 1989 = Bd. 28, S. 36–66.

Peterson, Robert: On the Use of College Students in Social Science Research: Insights from a Second-Order Meta-analysis, Journal of Consumer Research 2001 = Bd. 28, S. 450–461.

Pfeffer, Jeffrey: Size and Composition of Corporate Boards of Directors: The Organization and its Environment, Administrative Science Quarterly 1972 = Bd. 17, S. 218–228.

Pfeffer, Jeffrey: The Ombudsman: Evidence and Executive Compensation—Like Trains Passing in the Night? An Introduction to “Are Top Executives Paid Enough?”, Interfaces 2013 = Bd. 43, S. 578–579.

Pfeffer, Jeffrey & Robert Sutton: Evidence-Based Management, Harvard Business Review 2006 = Bd. 84, S. 62–74.

Pfeffer, Jeffrey & Robert Sutton: Hard Facts, Dangerous Half-Truths, and Total Nonsense: Profiting from Evidence-Based Management, Harvard 2006.

Pflüger, Almut: Rechtstatsachenforschung in der Praxis. Anwendungsbeispiele für die Vielfalt der Forschungsmöglichkeiten – vom Aktenstudium bis zu repräsentativen Umfragen, in: Brand/Strempel, Soziologie des Rechts. Festschrift für Erhard Blankenburg zum 60. Geburtstag, Baden-Baden 1998, S. 561–568.

Pflüger, Almut: Fragen über Fragen: Aktuelles aus der Rechtsdemoskopie, Gewerblicher Rechtsschutz und Urheberrecht. Praxis im Immaterialgüter- und Wettbewerbsrecht 2011, S. 51–54.

Pollman, Elizabeth: Strengthening Special Committees, UC Davis Business Law Journal 2009 = Bd. 9, S. 137–170.

Poole, Marshall Scott, Andrea Hollingshead, Joseph McGrath, Richard Moreland & John Rohrbaugh: Interdisciplinary Perspectives on Small Groups, in: Poole/Hollingshead, Theories of Small Groups: Interdisciplinary Perspectives, Thousand Oaks 2005, S. 1–20.

Popper, Karl: Alles Leben ist Problemlösen: Über Erkenntnis, Geschichte und Politik, München 2001.

Posner, Richard: The State of Legal Scholarship Today: A Comment on Schlag, The Georgetown Law Journal 2009 = Bd. 97, S. 845–855.

Posner, Richard: Judicial Opinions and Appellate Advocacy in Federal Courts – One Judge’s Views, Duquesne Law Review 2013 = Bd. 51, S. 3–39.

Priester, Hans-Joachim: Stichentscheid bei zweiköpfigem Vorstand, Die Aktiengesellschaft 1984, S. 253.

Prigge, Stefan: A Survey of German Corporate Governance, in: Hopt/Kanda/Roe/Wymeersch/Prigge, Comparative Corporate Governance. The State of the Art and Emerging Research, Oxford 1998, S. 943–1044.

Prince, Mary Miles: The Bluebook: A Uniform System of Citation, 19. Aufl., Cambridge 2010.

Prucher, Jeff: Brave New Words: The Oxford Dictionary of Science Fiction, Oxford 2007.

Puncochar, Judith & Paul Fox: Confidence in Individual and Group Decision Making: When “Two Heads” Are Worse Than One, Journal of Educational Psychology 2004 = Bd. 96, S. 582–591.

Quatember, Andreas: Statistik ohne Angst vor Formeln, 3. Aufl., München 2011.

Rabl, Tanja: Do contextual factors matter? An investigation of ethical judgments of corrupt acts, Zeitschrift für Betriebswirtschaft 2012 = Bd. 82, S. 5–32.

Rachlinski, Jeffrey: The “New” Law and Psychology: A Reply to Critics, Skeptics and Cautious Supporters, Cornell Law Review 1999 = Bd. 85, S. 739–766.
Rachlinski, Jeffrey: Evidence-Based Law, Cornell Law Review 2011 = Bd. 96, S. 901–923.
Rachlinski, Jeffrey: The Psychological Foundations of Behavioral Law and Economics, University of Illinois Law Review 2011, S. 1675–1696.
Radjuk, Anna: Das neue russische GmbH-Recht, Recht der internationalen Wirtschaft 2009, S. 592–597.
Radner, R.: Team Decision Problems, The Annals of Mathematical Statistics 1962 = Bd. 33, S. 857–881.
Raiser, Thomas: Das Unternehmen als Organisation. Kritik und Erneuerung der juristischen Unternehmenslehre, Habil., Berlin 1969.
Raiser, Thomas: Beiträge zur Rechtssoziologie, Baden-Baden 2011.
Ramaiya, Arunachala & Yeshwant Vishnu Chandrachud: Guide to the Companies Act, 16. Aufl., New Delhi 2006.
Rangachari, Patangi: Evidence-based medicine: old French wine with a new Canadian label?, Journal of the Royal Society of Medicine 1997 = Bd. 90, S. 280–284.
Rapp, Marc Steffen, Philipp Schaller & Michael Wolff: Aktienbasierte Langfristanreize im Rahmen der Vorstandsvergütung. Evidenz auf Basis Deutscher Prime Standard Unternehmen, Die Betriebswirtschaft 2011, S. 311–330.
Raspe, Heiner: Evidence based medicine: Modischer Unsinn, alter Wein in neuen Schläuchen oder aktuelle Notwendigkeit?, Zeitschrift für ärztliche Fortbildung 1996 = Bd. 90, S. 553–562.
Raspe, Heiner: Der „allgemein anerkannte Stand der medizinischen Erkenntnisse“ – das Konzept der Evidence-Based Medicine, Gesundheitsrecht 2011, S. 449–454.
Raspe, Heiner: Von der Erfahrung zur Evidenz. Zum Wandel der Wissensgrundlagen in der Medizin, Gesundheitsrecht 2012, S. 584–591.
Rauscher, Thomas (Hg.): Münchener Kommentar zum FamFG, 2. Aufl., München 2013.
Read, Daniel: Monetary incentives, what are they good for?, Journal of Economic Methodology 2005 = Bd. 12, S. 265–276.
Redeker, Konrad: Von der Unsitte des Schreibens in eigener Sache, Neue Juristische Wochenschrift 1983, S. 1034–1035.
Rehbinder, Manfred: Arthur Nußbaum [Geleitwort], in: Nußbaum, Die Rechtstatsachenforschung. Programmschriften und praktische Beispiele, Berlin 1968, S. 9–17.
Rehbinder, Manfred: The Development and Present State of Fact Research in Law in the United States, Journal of Legal Education 1972, S. 567–589.
Reich-Rohrwig, Johannes: Verbreitung und Gesellschafterstruktur der GmbH in Österreich, in: Doralt/Nowotny, Kontinuität und Wandel. Beiträge zum Unternehmensrecht. Festschrift für Walther Kastner zum 90. Geburtstag, Wien 1992, S. 371–382.
Reich-Rohrwig, Johannes: Empirische Untersuchung über die GmbH in Österreich, in: Enzinger/Hügel/Dillenz, Aktuelle Probleme des Unternehmensrechts. Festschrift Gerhard Frotz zum 65. Geburtstag, Wien 1993, S. 381–400.
Reichardt, Charles: Criticisms of and an alternative to the Shadish, Cook, and Campbell validity typology, New Directions for Evaluation 2011, S. 43–53.
Reichsgericht: Urt. v. 3.1.1884 – Az. 2635/83, Entscheidungen des Reichsgerichts in Strafsachen 1884 = Bd. 9, S. 370–380.
Reichsgericht: Urt. v. 18.10.1917 – Az. VI 143/17, Entscheidungen des Reichsgerichts in Zivilsachen 1917 = Bd. 91, S. 72–80.

Reichsgericht: Urt. v. 4.10.1918 – Az. Rep. II. 498/17, Entscheidungen des Reichsgerichts in Zivilsachen 1918 = Bd. 93, S. 338–341.
Reichsgericht: Urt. v. 3.2.1920 – Az. II 272/19, Entscheidungen des Reichsgerichts in Zivilsachen 1920 = Bd. 98, S. 98–101.
Reilly, Brendan: The Essence Of EBM, British Medical Journal 2004 = Bd. 329, S. 991–992.
Reimer, Torsten, Andrea Reimer & Uwe Czienskowski: Decision-Making Groups Attenuate the Discussion Bias in Favor of Shared Information: A Meta-Analysis, Communication Monographs 2010 = Bd. 77, S. 121–142.
Reiner, Günter: Geheimsache Gesellschaftssatzung am Finanzmarkt Deutschland? Rechtliche Bedeutung, Publizitätspflichten und tatsächliche Verfügbarkeit für den Aktionär, Die Aktiengesellschaft 2006, S. 93–105.
Remus, William: Graduate students as surrogates for managers in experiments on business decision making, Journal of Business Research 1986 = Bd. 14, S. 19–25.
Remus, William: Will behavioral research on managerial decision making generalize to managers?, Managerial and Decision Economics 1996 = Bd. 17, S. 93–101.
Resch, Reinhard: Zur Ressortverteilung im Vorstand der Aktiengesellschaft, Der Gesellschafter. Zeitschrift für Gesellschafts- und Unternehmensrecht 2000, S. 2–6.
Rhode, Deborah: Legal Scholarship, Harvard Law Review 2002 = Bd. 115, S. 1327–1361.
Richard, Dan, Charles Jr. Bond & Juli Stokes-Zoota: One Hundred Years of Social Psychology Quantitatively Described, Review of General Psychology 2003 = Bd. 7, S. 331–363.
Rieger, Harald: Gesetzeswortlaut und Rechtswirklichkeit im Aktiengesetz, in: Lutter/Scholz/Sigle, Festschrift für Martin Peltzer zum 70. Geburtstag, Köln 2001, S. 339–357.
Riegert, Robert: Empirical Research about Law: The German Picture, with Comparisons and Observations, Dickinson International Law Annual 1983 = Bd. 2, S. 1–64.
Rieß, Peter: Der Niedergang des Kollegialprinzips in der Strafgerichtsbarkeit, in: Jung/Luxenburger/Wahle, Festschrift für Egon Müller, Baden-Baden 2008, S. 599–610.
Ring, Gerhard & Line Olsen-Ring: Einführung in das skandinavische Recht, München 1999.
Ringleb, Henrik-Michael, Thomas Kremer, Marcus Lutter & Axel von Werder: Die Kodex-Änderungen vom Mai 2010, Neue Zeitschrift für Gesellschaftsrecht 2010, S. 1161–1166.
Ringleb, Henrik-Michael, Thomas Kremer, Marcus Lutter & Axel von Werder (Hg.): Kommentar zum Deutschen Corporate Governance Kodex (Kodex-Kommentar), 5. Aufl., München 2014.
Ritchie, Stuart, Richard Wiseman & Christopher French: Failing the Future: Three Unsuccessful Attempts to Replicate Bem's 'Retroactive Facilitation of Recall' Effect, PLoS ONE 2012 e33423, Zugriff am 27.6.2014 unter http://dx.doi.org/10.1371%2F journal.pone.0033423.
Rogers, Tom: Type I and Type II Errors – Making Mistakes in the Justice System, Amazing Applications of Probability and Statistics, Zugriff am 11.9.2012 unter http://www.intuitor.com/statistics/T1T2Errors.html.
Röhl, Klaus: Rechtssoziologie. Ein Lehrbuch, Köln 1987.
Röhl, Klaus: Franz von Benda-Beckmann (1941-2013), RSOZBLOG.de, Zugriff am 16.2.2014 unter http://www.rsozblog.de/franz-von-benda-beckmann-1941-2013/.

Roll, Richard: The Hubris Hypothesis of Corporate Takeovers, Journal of Business 1986 = Bd. 59, S. 197–216.

Rosenberg, William & Anna Donald: Evidence based medicine: an approach to clinical problem-solving, British Medical Journal 1995 = Bd. 310, S. 1122–1126.

Rosenblueth, Arturo & Norbert Wiener: The Role of Models in Science, Philosophy of Science 1945 = Bd. 12, S. 316–321.

Rosenthal, Robert: The file drawer problem and tolerance for null results, Psychological Bulletin 1979 = Bd. 86, S. 638–641.

Rosenthal, Robert & Robin DiMatteo: Meta-Analysis: Recent Developments in Quantitative Methods for Literature Reviews, Annual Review of Psychology 2001 = Bd. 52, S. 59–82.

Rossi, Peter Henry & Steven Nock: Measuring Social Judgments: The Factorial Survey Approach, Beverly Hills 1982.

Rost, Katja & Margit Osterloh: Managementmode Pay-for-Performance, Universität Zürich, Professur für Organisation, Technologie- und Innovationsmanagement, Zugriff am 27.6.2014 unter http://www.uzh.ch/iou/orga/ssl-dir/wiki/uploads/Main/Rost_Osterloh_2007.pdf.

Rost, Katja & Margit Osterloh: Management Fashion Pay-for-Performance for CEOs, Schmalenbach Business Review 2009 = Bd. 61, S. 119–149.

Roth Pellanda, Katja: Organisation des Verwaltungsrates. Zusammensetzung, Arbeitsteilung, Information und Verantwortlichkeit, Diss. 2007, Zürich 2007.

Rothwell, Peter & Christopher Martyn: Reproducibility of peer review in clinical neuroscience: Is agreement between reviewers any greater than would be expected by chance alone?, Brain 2000 = Bd. 123, S. 1964–1969.

Rottnauer, Achim: Konstituierung der HV durch einen „unterbesetzten Vorstand”. Kommentar zu OLG Dresden, NZG 2000, 426 („Sachsenmilch”), Neue Zeitschrift für Gesellschaftsrecht 2000, S. 414–418.

Rousseau, Denise: Organizational Behavior in the New Organizational Era, Annual Review of Psychology 1997 = Bd. 48, S. 515–546.

Rousseau, Denise: Is There Such A Thing As "Evidence-Based Management"?, Academy of Management Review 2006 = Bd. 31, S. 256–269.

Rowe, Gene & George Wright: Differences in Expert and Lay Judgments of Risk: Myth or Reality?, Risk Analysis 2001 = Bd. 21, S. 341–356.

Rüb, Friedbert & Holger Straßheim: Politische Evidenz – Rechtfertigung durch Verobjektivierung?, in: Geis/Nullmeier/Daase, Der Aufstieg der Legitimitätspolitik. Rechtfertigung und Kritik politisch-ökonomischer Ordnungen (Leviathan Sonderbd. 27), Baden-Baden 2012, S. 377–397.

Rubinstein, Ariel: A Sceptic's Comment on the Study of Economics, The Economic Journal 2006 = Bd. 116, S. C1–C9.

Rudolph, George: Further Thoughts on the One and Two Director Statutes, The Business Lawyer 1965 = Bd. 20, S. 781–788.

Rustige, Marc & Michael Grote: Der Einfluss von Diversifikationsstrategien auf den Aktienkurs deutscher Unternehmen, Zeitschrift für betriebswirtschaftliche Forschung 2009, S. 470–498.

Rzepka, Walter: Gegen den verstärkten Einsatz von Einzelrichtern im Verwaltungsprozess, Bayerische Verwaltungsblätter 1991, S. 460–461.

Säcker, Franz Jürgen (Hg.): Münchener Kommentar zum Bürgerlichen Gesetzbuch Bd. 1: §§ 1-240 BGB, ProstG, AGG, 6. Aufl., München 2012.

Sackett, David, William Rosenberg, Muir Gray, Brian Haynes & Scott Richardson: Evidence Based Medicine: What It Is And What It Isn’t, British Medical Journal 1996 = Bd. 312, S. 71–72.
Saenger, Ingo (Hg.): Zivilprozessordnung, FamFG, Europäisches Verfahrensrecht. Handkommentar, 5. Aufl., Baden-Baden 2013.
Saks, Michael: On Tapp (And Levine), Michigan Law Review 1979 = Bd. 77, S. 892–898.
Saks, Michael & Mollie Weighner Marti: A Meta-Analysis of the Effects of Jury Size, Law and Human Behavior 1997 = Bd. 21, S. 451–467.
Salsburg, David: The Lady Tasting Tea: How Statistics Revolutionized Science in the Twentieth Century, New York 2001.
Sanderson, Ian: Evaluation, Policy Learning and Evidence-Based Policy Making, Public Administration 2002 = Bd. 80, S. 1–22.
Sauter, Eugen (Begr.), Gerhard Schweyer & Wolfram Waldner (Hg.): Der eingetragene Verein. Gemeinverständliche Erläuterung des Vereinsrechts unter Berücksichtigung neuester Rechtsprechung, 19. Aufl., München 2010.
Sawicki, Peter: Evidenzbasierte Medizin: Keine „Zwangsjacke“ für den Arzt, Deutsches Ärzteblatt 2005 = Bd. 102, S. A 888–892.
Schäfer, Hans-Bernd & Claus Ott: Lehrbuch der ökonomischen Analyse des Zivilrechts, 5. Aufl., Berlin 2012.
Scharff, M.M.: Understanding WorldCom’s Accounting Fraud: Did Groupthink Play a Role?, Journal of Leadership & Organizational Studies 2005 = Bd. 11, S. 109–118.
Scheuerle, Wilhelm: Juristische Evidenzen, Zeitschrift für Zivilprozeß 1971 = Bd. 84, S. 241–297.
Schießl, Harald & Jan Küpperfahrenberg: Steuerrechtliche Haftung der Vorstände von Vereinen und Verbänden – Risiko, Vermeidungsstrategie, Versicherbarkeit, Deutsches Steuerrecht 2006, S. 445–450.
Schlegelberger, Franz, Leo Quassowski, Gustav Herbig, Ernst Geßler & Wolfgang Hefermehl: Aktiengesetz vom 30. Januar 1937, 2. Aufl., Berlin 1937.
Schleicher, Matthias: Vorstand, Vorstandsmitglieder und Aufsichtsräte von Aktiengesellschaften koreanischen Rechts, Diss. 2006, Zürich 2007.
Schlueter, Elmar & Peter Schmidt: Special issue: Survey experiments, Methodology: European Journal of Research Methods for the Behavioral and Social Sciences 2010 = Bd. 6, S. 93–95.
Schmidt, Frank-Hermann: Verhaltensforschung und Recht. Ethologische Materialien zu einer Rechtsanthropologie, Diss. 1981, Berlin 1982.
Schmidt, Johann: Die Besetzung der Richterbank (II), in: Zehn Jahre Verwaltungsgerichtsordnung. Bewährung und Reform. Vorträge und Diskussionsbeiträge der 38. Staatswissenschaftlichen Fortbildungstagung der Hochschule für Verwaltungswissenschaften Speyer, Berlin 1970, S. 85–95.
Schmitt, Hermann & Thorsten Vogt: Stärkung der Rechte von Aktionären – Reform des russischen Aktiengesetzes, Recht der internationalen Wirtschaft 2002, S. 762–767.
Schneider, Carl & Lee Teitelbaum: Life’s Golden Tree: Empirical Scholarship and American Law, Utah Law Review 2006, S. 53–106.
Schneider, Michael: Die Beschlussfähigkeit und Beschlussfassung von Kollegialorganen, Diss., Bochum 2000.
Schneider, Uwe: Die Wahrnehmung öffentlich-rechtlicher Pflichten durch den Geschäftsführer. Zum Grundsatz der Gesamtverantwortung bei mehrköpfiger Geschäftsführung in der konzernfreien GmbH und im Konzern, in: Lutter/Ulmer/Zöllner, Festschrift 100 Jahre GmbH-Gesetz, Köln 1992, S. 473–499.

Schnieders, Ferit: Haftungsfreiräume für unternehmerische Entscheidungen in Deutschland und Italien, Diss. 2007, Köln 2009.

Scholl, Claus: Sicherheit und Wahrscheinlichkeit – statistische, medizinische und juristische Aspekte, Neue Juristische Wochenschrift 1983, S. 319–320.

Schön, Wolfgang: Quellenforscher und Pragmatiker – Ein Schlusswort, in: Engel/Schön, Das Proprium der Rechtswissenschaft, Tübingen 2007, S. 313–321.

Schooler, Jonathan: Unpublished results hide the decline effect, Nature 2011 = Bd. 470, S. 437–437.

Schreiner, Helmut: Evidenz und Rechtsmethodologie, in: Krawietz/Mayer-Maly/Weinberger, Objektivierung des Rechtsdenkens. Gedächtnisschrift für Ilmar Tammelo, Berlin 1984, S. 543–555.

Schremper, Ralf: Kapitalmarktrelevanz deutscher Aktienrückkaufprogramme, Zeitschrift für betriebswirtschaftliche Forschung 2003 = Bd. 55, S. 578–605.

Schubert, Werner: Die Entstehung des Aktiengesetzes vom 18. Juli 1884, in: Schubert/Hommelhoff, Hundert Jahre modernes Aktienrecht. Eine Sammlung von Texten und Quellen zur Aktienrechtsreform 1884 mit zwei Einführungen, Berlin 1985, S. 1–52.

Schuck, Peter: Why Don't Law Professors Do More Empirical Research?, Journal of Legal Education 1989 = Bd. 39, S. 323–336.

Schüller, Annette: Vorstandsvergütung. Gesellschaftsrechtliche Fragen der Vergütung des Vorstands in der börsennotierten Aktiengesellschaft, Diss. 2000, Baden-Baden 2002.

Schulz Von Thun: Miteinander reden Bd. 1: Störungen und Klärungen. Allgemeine Psychologie der Kommunikation, 44. Aufl., Rowohlt Taschenbuch 2006.

Schweizer, Robert: Empirische Rechtsforschung, in: GfK e.V., Jahrbuch der Absatz- und Verbrauchsforschung, Berlin 1976, S. 386–429.

Schweizer, Robert: Die Entdeckung der pluralistischen Wirklichkeit: Durchschnittsleser, Presserecht, Verständiger Verbraucher, Wettbewerbsrecht, Wertvorstellungen, Grundnorm, 3. Aufl., Berlin 2000.

Schwenk, Charles: Effects of devil's advocacy and dialectical inquiry on decision making: A meta-analysis, Organizational Behavior and Human Decision Processes 1990 = Bd. 47, S. 161–176.

Sears, David: College sophomores in the laboratory: Influences of a narrow data base on social psychology's view of human nature, Journal of Personality and Social Psychology 1986 = Bd. 51, S. 515–530.

Sedlmeier, Peter & Frank Renkewitz: Forschungsmethoden und Statistik in der Psychologie, München 2008.

Seidenfeld, Mark: Cognitive Loafing, Social Conformity, and Judicial Review of Agency Rulemaking, Cornell Law Review 2002 = Bd. 87, S. 486–548.

Selten, Reinhard: Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperiments, in: Sauermann, Beiträge zur experimentellen Wirtschaftsforschung, Tübingen 1967, S. 136–168.

Sendler, Horst: Zum Schreiben in eigener Sache, in: Rüthers/Stern, Freiheit und Verantwortung im Verfassungsstaat. Festgabe zum 10-jährigen Jubiläum der Gesellschaft für Rechtspolitik, München 1984, S. 413–432

Senn, Stephen: Bayesian, Likelihood, and Frequentist Approaches to Statistics. A comparison of methods, Applied Clinical Trials 2003, S. 35–38.

Senn, Stephen: Safety first?, Significance 2007, S. 79–80.

Seth, Anju, Kevin Carlson, Donald Hatfield & Hung-Wen Lan: So what? Beyond statistical significance to substantive significance in strategy research, in:

Bergh/Ketchen, Research Methodology in Strategy and Management, Bradford 2009, S. 3–27.

Shadish, William, Thomas Cook & Donald Campbell: Experimental and Quasi-Experimental Designs for Generalized Causal Inference, Belmont 2002.

Sharfman, Bernard & Steven Toll: A Team Production Approach to Corporate Law and Board Composition, Northwestern University Law Review Colloquy 2009 = Bd. 103, S. 380–392.

Sherman, Lawrence: Evidence-Based Policing, Police Foundation: Ideas in American Policing, Zugriff am 15.12.2012 unter www.policefoundation.org/sites/pftest1.drupalgardens.com/files/Sherman%20(1998)%20-%20Evidence-Based%20Policing.pdf.

Sherman, Lawrence: Evidence-Based Crime Prevention, London 2002.

Sherman, Lawrence: Evidence-Based Policing: Social Organization of Information for Social Control, in: Waring/Weisburd, Crime & Social Organisation, New Brunswick 2002, S. 217–248.

Shleifer, Andrei & Robert Vishny: A Survey of Corporate Governance, The Journal of Finance 1997 = Bd. 52, S. 737–783.

Sibert, Anne: Central Banking by Committee, International Finance 2006 = Bd. 9, S. 145–168.

Simmons, Joseph, Leif Nelson & Uri Simonsohn: False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant, Psychological Science 2011 = Bd. 22, S. 1359–1366.

Simon, Mark & Susan Houghton: The Relationship between Overconfidence and the Introduction of Risky Products: Evidence from a Field Study, The Academy of Management Journal 2003 = Bd. 46, S. 139–149.

Sims, Ronald: Linking groupthink to unethical behavior in organizations, Journal of Business Ethics 1992 = Bd. 11, S. 651–662.

Skog, Rolf & Catarina Fäger: The Swedish Companies Act: An Introduction, Stockholm 2007.

Slavin, Robert: Evidence-Based Education Policies: Transforming Educational Practice and Research, Educational Researcher 2002 = Bd. 31, S. 15–21.

Smith, Adrian: Mad Cows and Ecstasy: Chance and Choice in an Evidence-Based Society, Journal of the Royal Statistical Society. Series A (Statistics in Society) 1996 = Bd. 159, S. 367–383.

Smith, James & Thomas Kida: Heuristics and biases: Expertise and task realism in auditing, Psychological Bulletin 1991 = Bd. 109, S. 472–489.

Smith, Laurence, Lisa Best, Alan Stubbs, Andrea Bastiani Archibald & Roxann Roberson-Nay: Constructing knowledge: The role of graphs and tables in hard and soft psychology, American Psychologist 2002 = Bd. 57, S. 749–761.

Snellman, Hannes: The Finnish Companies Act of 1978, Helsinki 1984.

Sniezek, Janet: Groups under uncertainty: An examination of confidence in group decision making, Organizational Behavior and Human Decision Processes 1992 = Bd. 52, S. 124–155.

Sohst, Wolfgang: Das spanische Aktien- und GmbH-Gesetz, 1. Aufl., Berlin 2002.

Sohst, Wolfgang: Spanisches Gesellschaftsrecht (Textsammlung), 5. Aufl., Berlin 2012.

Son, Chu-ch'an: Das koreanische Aktienrecht, 2. Aufl., Neuwied 1989.

Sonnenberger, Hans Jürgen & Christian Autexier: Einführung in das französische Recht, 3. Aufl., Heidelberg 2000.

Spiel, Christiane: Evidenzbasierte Bildungspolitik und Bildungspraxis – eine Fiktion? Problemaufriss, Thesen, Anregungen, Psychologische Rundschau 2009 = Bd. 60, S. 255–256.

Spindler, Gerald & Eberhard Stilz (Hg.): Kommentar zum Aktiengesetz, 2 Bde., 2. Aufl., München 2010.

Spoerri, James: One Incorporator. One Director, The Business Lawyer 1963 = Bd. 19, S. 305–309.

Springer, Michael: Komplexität und Emergenz, Spektrum der Wissenschaft 2012, S. 48–54.

Stallberg, Christian: Evidenz-basierte Medizin als Rechtsbegriff – Funktion, Inhalt und Grenzen, Pharmarecht 2010, S. 5–12.

Stanley, Tom: Beyond Publication Bias, Journal of Economic Surveys 2005 = Bd. 19, S. 309–345.

Stanovich, Keith: How To Think Straight About Psychology, 5. Aufl., New York 1998.

Stanovich, Keith: How To Think Straight About Psychology, 10. Aufl., Boston 2012.

Stanton, Jeffrey: Galton, Pearson, and the Peas: A Brief History of Linear Regression for Statistics Instructors Journal of Statistics Education, Zugriff am 10.7.2012 unter http://www.amstat.org/publications/jse/v9n3/stanton.html.

Starck, Christian: Empirie in der Rechtsdogmatik, JuristenZeitung 1972, S. 609–614.

Starck, Christian: Der demokratische Verfassungsstaat: Gestalt, Grundlagen, Gefährdungen, Tübingen 1995.

Stasser, Garold, Dennis Stewart & Gwen Wittenbaum: Expert Roles and Information Exchange during Discussion: The Importance of Knowing Who Knows What, Journal of Experimental Social Psychology 1995 = Bd. 31, S. 244–265.

Stasser, Garold & William Titus: Pooling of unshared information in group decision making: Biased information sampling during discussion, Journal of Personality and Social Psychology 1985 = Bd. 48, S. 1467–1478.

Stasser, Garold & William Titus: Hidden Profiles: A Brief History, Psychological Inquiry 2003 = Bd. 14, S. 304–313.

Stasson, Mark & Scott Bradshaw: Explanations of Individual-Group Performance Differences, Small Group Research 1995 = Bd. 26, S. 296–308.

Staub, Hermann: Kommentar zum Handelsgesetzbuch, 14. Aufl. , Berlin 1933.

Steblay, Nancy, Harmon Hosch, Scott Culhane & Adam McWethy: The Impact on Juror Verdicts of Judicial Instruction to Disregard Inadmissible Evidence: A Meta-Analysis, Law and Human Behavior 2006 = Bd. 30, S. 469–492.

Steblay, Nancy Mehrkens, Jasmina Besirevic, Solomon Fulero & Belia Jimenez-Lorente: The Effects of Pretrial Publicity on Juror Verdicts, Law and Human Behavior 1999 = Bd. 23, S. 219–235.

Steckel-Berger, Gabriele: Statistische Datenanalyse für JuristInnen, in: Barta/Ganner/Lichtmannegger, Rechtstatsachenforschung – Heute. Tagungsband 2008, Innsbruck 2009, S. 185–204.

Stein, Rosemarie: Evidence based Medicine: Unsinn eliminieren, Deutsches Ärzteblatt 1998 = Bd. 95, S. A 862–864.

Steiner, Ivan: Group process and productivity, New York 1972.

Steiner, Peter & Christiane Atzmüller: Experimentelle Vignettendesigns in faktoriellen Surveys, Kölner Zeitschrift für Soziologie und Sozialpsychologie 2006 = Bd. 58, S. 117–146.

Steinitzer, Erwin: Ökonomische Theorie der Aktiengesellschaft, Leipzig 1908.

Stelkens, Paul: Das Gesetz zur Neuregelung des verwaltungsgerichtlichen Verfahrens (4. VwGOÄndG) das Ende einer Reform?, Neue Zeitschrift für Verwaltungsrecht 1991, S. 209–219.

Stenzel, Roman: Rechtliche und empirische Aspekte der Vorstandsvergütung, Diss. 2011, Jena 2012.

Sterling, Theodor, Wilf Rosenbaum & James Weinkam: Publication Decisions Revisited: The Effect of the Outcome of Statistical Tests on the Decision to Publish and Vice Versa, The American Statistician 1995 = Bd. 49, S. 108–112.

Sterling, Theodore: Publication Decisions and Their Possible Effects on Inferences Drawn from Tests of Significance–Or Vice Versa, Journal of the American Statistical Association 1959 = Bd. 54, S. 30–34.

Stigler, George: Economics: The Imperial Science?, The Scandinavian Journal of Economics 1984 = Bd. 86, S. 301–313.

Stigler, Stephen: Stigler's Law of Eponymy, Transactions of the New York Academy of Sciences 1980 = Bd. 39, S. 147–157.

Straus, Sharon & Giselle Jones: What Has Evidence Based Medicine Done For Us?, British Medical Journal 2004 = Bd. 329, S. 987–988.

Straus, Sharon & Finlay McAlister: Evidence-based medicine: a commentary on common criticisms, Canadian Medical Association Journal 2000 = Bd. 163, S. 837–841.

Strempel, Dieter: Empirische Rechtsforschung. Entwicklung und Beitrag für die Rechtspolitik, in: Broda/Deutsch/Schreiber/Vogel, Festschrift für Rudolf Wassermann zum sechzigsten Geburtstag, Neuwied 1985, S. 223–233.

Strempel, Dieter: Der Beitrag der empirischen Rechtsforschung zu einem realistischen Umgang mit Recht, in: Schäffer, Gesetzgebung und Rechtskultur. Internationales Symposion Salzburg 1986, Wien 1987, S. 87–93.

Strine, Leo: The Inescapably Empirical Foundation of the Common Law of Corporations, Delaware Journal of Corporate Law 2002 = Bd. 27, S. 499–518.

Struck, Gerhard: Gesetz und Chaos in Naturwissenschaft und Rechtswissenschaft, Juristische Schulung 1993, S. 992–999.

Suchman, Mark & Elizabeth Mertz: Toward a New Legal Empiricism: Empirical Legal Studies and New Legal Realism, Annual Review of Law and Social Science 2010 = Bd. 6, S. 555–579.

Sullivan, Ryan, Allan Timmermann & Halbert White: Dangers of data mining: The case of calendar effects in stock returns, Journal of Econometrics 2001 = Bd. 105, S. 249–286.

Sünner, Eckart: Diversity bei den Organen einer Aktiengesellschaft, Corporate Compliance Zeitschrift 2009, S. 185–190.

Sunstein, Cass: Deliberative Trouble? Why Groups Go to Extremes, The Yale Law Journal 2000 = Bd. 110, S. 71–119.

Sunstein, Cass: The Law of Group Polarization, Journal of Political Philosophy 2002 = Bd. 10, S. 175–195.

Sunstein, Cass: Group Judgments: Statistical Means, Deliberation, and Information Markets, New York University Law Review 2005 = Bd. 80, S. 962–1049.

Süß, Rembert & Thomas Wachter (Hg.): Handbuch des internationalen GmbH-Rechts, 2. Aufl., Bonn 2011.

Sutter, Matthias: Individual Behavior and Group Membership: Comment, American Economic Review 2009 = Bd. 99, S. 2247–2257.

Swarup, Jagadish: The Companies Act, 1956, Lucknow 1966.

Sweeney, Laura & Craig Haney: The influence of race on sentencing: A meta-analytic review of experimental studies, Behavioral Sciences & the Law 1992 = Bd. 10, S. 179–195.

Swieringa, Robert & Karl Weick: An Assessment of Laboratory Experiments in Accounting, Journal of Accounting Research 1982 = Bd. 20, S. S56–S101.

Takahashi, Eiji: Japanische Corporate Governance unter dem Gesellschaftsgesetz von 2005, Die Aktiengesellschaft 2007, S. 476–480.

Talley, Eric & Drew O'Kane: The Measure of a MAC: A Machine-Learning Protocol for Analyzing Force Majeure Clauses in MA Agreements, Journal of Institutional and Theoretical Economics 2012 = Bd. 168, S. 181–201.

Thamm, Robert: Die Organisationsautonomie der monistischen Societas Europaea bezüglich ihrer geschäftsführenden Direktoren, Neue Zeitschrift für Gesellschaftsrecht 2008, S. 132–134.

Thamm, Robert: Die rechtliche Verfassung des Vorstands der AG, Diss. 2007, Berlin 2008.

Thiele, Carmen: Regeln und Verfahren der Entscheidungsfindung innerhalb von Staaten und Staatenverbindungen. Staats- und kommunalrechtliche sowie europa- und völkerrechtliche Untersuchungen, Habil. 2007, Heidelberg 2008.

Thieme, Werner: Vom Publizieren in der Sache des Mandanten, Neue Juristische Wochenschrift 1983, S. 2015.

Thomson Reuters: Web of Science: Journal Title Abbreviations, ISI Web of Knowledge, Zugriff am 8.11.2012 unter http://images.webofknowledge.com/WOK46/help/WOS/A_abrvjt.html.

Thüsing, Gregor: ECLR – Auf der Suche nach dem iustum pretium der Vorstandstätigkeit. Überlegungen zur Angemessenheit im Sinne des § 87 Abs. 1 Satz 1 AktG, Zeitschrift für Unternehmens- und Gesellschaftsrecht 2003, S. 457–507.

Timmer, Antje: Plädoyer für die evidenzbasierte Medizin, Deutsches Ärzteblatt 2012 = Bd. 109, S. A 1418–1422.

Tindale, Scott & James Larson: Assembly bonus effect or typical group performance? A comment on Michaelsen, Watson, and Black (1989), Journal of Applied Psychology 1992 = Bd. 77, S. 102–105.

Tindale, Scott & James Larson: It's not how you frame the question, it's how you interpret the results, Journal of Applied Psychology 1992 = Bd. 77, S. 109–110.

Tirole, Jean: Hierarchies and Bureaucracies: On the Role of Collusion in Organizations, Journal of Law, Economics & Organization 1986 = Bd. 2, S. 181–214.

Tobin, James: Estimation of Relationships for Limited Dependent Variables, Econometrica 1958 = Bd. 26, S. 24–36.

Tomasic, Roman, Stephen Bottomley & Rob McQueen: Corporations Law in Australia, 2. Aufl., Sydney 2002.

Tontrup, Stephan: Zum unterschiedlichen Verhältnis der juristischen Teilfächer zu den Sozialwissenschaften. Kommentar zu Günther Jakobs und Joachim Schulz, in: Engel/Schön, Das Proprium der Rechtswissenschaft, Tübingen 2007, S. 192–202.

Tosi, Henry Jr. & Luis Gomez-Mejia: The Decoupling of CEO Pay and Performance: An Agency Theory Perspective, Administrative Science Quarterly 1989 = Bd. 34, S. 169–189.

Trevelyan, Rose: Optimism, overconfidence and entrepreneurial activity, Management Decision 2008 = Bd. 46, S. 986–1001.

Triebe, Benjamin: Der Einfluss des Kalenders auf die Kurse, Neue Zürcher Zeitung 9.8.2010, Zugriff am 27.6.2014 unter http://www.nzz.ch/aktuell/startseite/der-einfluss-des-kalenders-auf-die-kurse-1.7142949.

Trottier, Kim & Irene Gordon: Students as Surrogates for Managers: Evaluating the Conclusions from a Replicated Accounting Experiment, Social Science Research Network, Zugriff am 29.8.2012 unter http://ssrn.com/paper=1928139.

Trujillo, Bernard: Patterns in a Complex System: An Empirical Study Of Valuation in Business Bankruptcy Cases, UCLA Law Review 2005 = Bd. 53, S. 357–404.

Tüxen, Markus: Kollegialprinzip oder Einzelrichter. Die Entwicklung des zivilrechtlichen Spruchkörpers bei den Landgerichten sowie den Oberlandesgerichten seit 1879 unter

besonderer Berücksichtigung der Reformen von 1974 und 1976, Diss. 2009, Frankfurt 2009.

Ueberwasser, Heinrich: Das Kollegialprinzip. Seine Grundsätze und Konkretisierungen im Bereiche von Regierung und Verwaltung unter besonderer Berücksichtigung des schweizerischen Bundesrates, Diss. 1988, Basel 1989.

Ule, Carl Hermann: Anselm Feuerbach heute. Die Bedeutung seiner Lehren für Gerichtsverfassung und gerichtliches Verfahren, Deutsches Verwaltungsblatt 1979, S. 797–807.

Ulen, Thomas: A Nobel Prize in Legal Science: Theory, Empirical Work, and the Scientific Method in the Study of Law, University of Illinois Law Review 2002, S. 875–920.

Ulmer, Peter: Zum Für und Wider des Publizieren von Rechtsgutachten, Neue Juristische Wochenschrift 1983, S. 2923–2924.

Universität Jena: Institut für Rechtstatsachenforschung » Publikationen » Aufsätze und Anmerkungen, Rechtswissenschaftliche Fakultät der Universität Jena, Zugriff am 3.8.2014 unter www.rewi.uni-jena.de/Fakult%C3%A4t/Institute/Institut+f%C3%BCr+Rechtstatsachenforschung/Publikationen/Aufs%C3%A4tze+und+Anmerkungen.html.

Upmeier, Arne: Fakten im Recht. Eine Untersuchung zur Tatsachenfeststellung im Rechtsprozess, Diss. 2006, Baden-Baden 2010.

van Aaken, Anne: „Rational Choice“ in der Rechtswissenschaft. Zum Stellenwert der ökonomischen Theorie im Recht, 1. Aufl., Baden-Baden 2003.

van Belle, Gerald: Statistical Rules of Thumb, 2. Aufl., Hoboken 2008.

Van Gestel, Rob: Evidence-based lawmaking and the quality of legislation. Regulatory impact assessments in the European Union and the Netherlands, in: Schäffer/Iliopoulos-Strangas, State modernization in Europe / Staatsmodernisierung in Europa. Tagungsband der Societas Iuris Publici Europaei (SIPE), Berlin 2007, S. 139–165.

Van Hulle, Karel & Harald Gesell: European Corporate Law, Baden-Baden 2006.

VanVoorhis, Carmen Wilson & Betsy Levonian Morgan: Statistical Rules of Thumb: What We Don’t Want to Forget About Sample Sizes, Psi Chi Journal of Undergraduate Research 2001 = Bd. 6, S. 139–141.

Verschiedene: Replication, Replication, Replication, The Psychologist 2012 = Bd. 25, S. 346–357.

Vickers, Andrew: What is a p-value anyway? 34 Stories To Help You Actually Understand Statistics, Boston 2010.

Vidmar, Neil & Valerie Hans: American Juries: The Verdict, Amherst 2007.

Viken, Monica: Legal Aspects Regarding the Use of Market Surveys as Evidence, Nordiskt Immateriellt Rättsskydd 2012 = Bd. 81, S. 220–234.

Vogel, Louis & Françoise Perochon: Code de commerce (Commercial Code), Legifrance translations, Zugriff am 8.10.2012 unter http://www.legifrance.gouv.fr/content/download/1951/13685/version/5/file/Code_32.pdf.

Vogel, Wolfgang: Aktienrecht und Aktienwirklichkeit – Organisation und Augabenteilung von Vorstand und Aufsichtsrat. Eine empirische Untersuchung deutscher Aktiengesellschaften, Diss. 1978, Baden-Baden 1980.

von Hein, Jan: Vom Vorstandsvorsitzenden zum CEO?, Zeitschrift für das gesamte Handels- und Wirtschaftsrecht 2002 = Bd. 166, S. 464–502.

von Hoffmann, Bernd & Gottfried von Waldheim: Das spanische Aktienrecht, 3. Aufl., Frankfurt/M. 1975.

von Jhering, Rudolph: Der Zweck im Recht. Erster Band, 4. Aufl., Leipzig 1904.

von Kirchmann, Julius: Die Werthlosigkeit der Jurisprudenz als Wissenschaft. Ein Vortrag, gehalten in der juristischen Gesellschaft zu Berlin, Berlin 1848.

Von Rohr, Alexandre: Evolutionsbiologische Grundlagen des Rechts. Zum Einfluss neurogenetischer Information auf das Recht. Ein Beitrag zur Rechtsethologie unter besonderer Berücksichtigung des Vertrauens im Recht, Diss. 2000, Berlin 2001.

von Rohr, Alexandre: Verhaltensforschung und Recht: zur Funktion von Normen in einer komplexen Welt, recht. Zeitschrift für juristische Ausbildung und Praxis 2004 = Bd. 22, S. 22–26.

Vorberg, Dirk & Sven Blankenberger: Die Auswahl statistischer Tests und Maße, Psychologische Rundschau 1999 = Bd. 50, S. 157–164.

Vranken, Jan: Exploring the Jurist’s Frame Of Mind. Constraints and Preconceptions in Civil Law Argumentation, Deventer 2006.

Wagner, Clifford: Simpson’s Paradox in Real Life, The American Statistician 1982 = Bd. 36, S. 46–48.

Wagner, John, J. L. Stimpert & Edward Fubara: Board Composition and Organizational Performance: Two Studies of Insider/Outsider Effects, Journal of Management Studies 1998 = Bd. 35, S. 655–677.

Wagner, Michael & Bernd Weiß: Meta-Analyse als Methode der Sozialforschung, Kölner Zeitschrift für Soziologie und Sozialpsychologie. Sonderheft 44: Methoden der Sozialforschung 2004, S. 479–504.

Wainer, Howard: The Most Dangerous Profession: A Note on Nonsampling Error, Psychological Methods 1999 = Bd. 4, S. 250–256.

Wainer, Howard & David Thissen: Graphical Data Analysis, Annual Review of Psychology 1981 = Bd. 32, S. 191–241.

Wainer, Howard & Paul Velleman: Statistical Graphics: Mapping the Pathways of Science, Annual Review of Psychology 2001 = Bd. 52, S. 305–335.

Wallander, Lisa: 25 years of factorial surveys in sociology: A review, Social Science Research 2009 = Bd. 38, S. 505–520.

Walshe, Kieran & Thomas Rundall: Evidence-based Management: From Theory to Practice in Health Care, Milbank Quarterly 2001 = Bd. 79, S. 429–457.

Walz, Christian: Das Ziel der Auslegung und die Rangfolge der Auslegungskriterien, Zeitschrift für das Juristische Studium 2010 = Bd. 3, S. 482–490.

Wang, Xia, Min Zhang & Fusheng Yu: Managerial overconfidence and over-investment: Empirical evidence from China, Frontiers of Business Research in China 2009 = Bd. 3, S. 453–469.

Wangler, Clemens: Fehlerquellen in der Unternehmensbewertung, Der Sachverständige – Fachzeitschrift für Sachverständige, Kammern, Gerichte und Behörden 2005, S. 245–249.

Weber-Rey, Daniela & Friederike Handt: Vielfalt/Diversity im Kodex – Selbstverpflichtung, Bemühenspflicht und Transparenz, Neue Zeitschrift für Gesellschaftsrecht 2011, S. 1–5.

Weber, Bernhard & Guido Hertel: Motivation gains of inferior group members: A meta-analytical review, Journal of Personality and Social Psychology 2007 = Bd. 93, S. 973–993.

Weber, James: Scenarios in Business Ethics Research: Review, Critical Assessment, and Recommendations, Business Ethics Quarterly 1992 = Bd. 2, S. 137–160.

Weber, Klaus: Vier Augen sehen mehr als zwei. Überlegungen zu Einzelrichterentscheidung und Kollegialprinzip, Zeitschrift für Rechtspolitik 1997, S. 134–137.

Weber, Max: Grundriss der Sozialökonomik III. Abteilung: Wirtschaft und Gesellschaft, Tübingen 1922.
Webster, Murray & Jane Sell (Hg.): Laboratory Experiments in the Social Sciences, Burlington 2007.
Wedde, Rainer: Neues im russischen Gmbh-Recht, osteuropa Recht 2009 = Bd. 55, S. 154–166.
Wegen, Gerhard: Gesellschaftsrecht in Indien: eine Einführung mit vergleichenden Tabellen, München 1997.
Weibel, Antoinette, Katja Rost & Margit Osterloh: Gewollte und ungewollte Anreizwirkungen von variablen Lohnen: Disziplinierung der Agenten oder Crowding-Out?, Zeitschrift für betriebswirtschaftliche Forschung 2007 = Bd. 59, S. 1029–1054.
Weibel, Antoinette, Katja Rost & Margit Osterloh: Pay for Performance in the Public Sector—Benefits and (Hidden) Costs, Journal of Public Administration Research and Theory 2010 = Bd. 20, S. 387–412.
Weigand, Christoph: Statistik mit und ohne Zufall. Eine anwendungsorientierte Einführung 2. Aufl., Heidelberg 2009.
Weiß, Bernd & Michael Wagner: The Identification and Prevention of Publication Bias in the Social Sciences and Economics, Jahrbücher für Nationalökonomie und Statistik 2011 = Bd. 231, S. 661–684.
Welling, Bruce: Corporate Law in Canada: The governing principles, 2. Aufl., Toronto 1991.
Welti, Felix & Heiner Raspe: Evidenz und Akzeptanz von Medizin und Recht, Neue Juristische Wochenschrift 2002, S. 874–875.
Werner, Winfried: Anm. zu BGH Urt. v. 23.10.1975 (Az. II ZR 90/73), Die Aktiengesellschaft 1976, S. 45–47.
Wessel, Maximilian, Henry Lahr & Olaf Ehrhardt: Kursreaktionen bei der Ankündigung von Squeeze-outs in Deutschland, M&A Review 2004, S. 361–369.
Westphal, James & Edward Zajac: Who Shall Govern? CEO/Board Power, Demographic Similarity, and New Director Selection, Administrative Science Quarterly 1995 = Bd. 40, S. 60–83.
Wettich, Carsten: Vorstandsorganisation in der Aktiengesellschaft – zugleich ein Beitrag zum Kollegialprinzip und dem Grundsatz der Gesamtverantwortung, Diss. 2008, Köln 2008.
Wheeler, Sally: Contracts and Corporations, in: Cane/Kritzer, The Oxford Handbook of Empirical Legal Research, Oxford 2010, S. 125–150.
Whyte, Glen: Escalating Commitment in Individual and Group Decision Making: A Prospect Theory Approach, Organizational Behavior and Human Decision Processes 1993 = Bd. 54, S. 430–455.
Wicke, Hartmut: Der CEO im Spannungsverhältnis zum Kollegialprinzip. Gestaltungsüberlegungen zur Leitungsstruktur der AG, Neue Juristische Wochenschrift 2007, S. 3755–3758.
Wiedemann, Herbert: Gesellschaftsrecht. Bd. 1: Grundlagen München 1980.
Wietek, Siegfried & Pierre Chomiac de Sas: Die kleine AG – die „SAS" – in Frankreich. Die optimale Rechtsform für französische Niederlassungen, flexibel – einfach – kostengünstig, Köln 2003.
Wilcox, Rand: Fundamentals of Modern Statistical Methods. Substantially Improving Power and Accuracy, 2. Aufl., New York 2010.
Williamson, Oliver: Corporate Boards of Directors: In Principle and in Practice, Journal of Law, Economics & Organization 2008 = Bd. 24, S. 247–272.

Windbichler, Christine: Zukunft des Gesellschaftsrechts: Orientierungen für die kapitalmarktorientierte Aktiengesellschaft in: Grundmann/Kloepfer/Paulus/Schröder/Werle, Festschrift 200 Jahre Juristische Fakultät der Humboldt-Universität zu Berlin, Berlin 2010, S. 1079–1098.

Windeler, Jürgen: Methodische Grundlagen einer evidenzbasierten Medizin, Gesundheitswesen 2008 = Bd. 70, S. 418–430.

Winer, Russell: Experimentation in the 21st Century: The Importance of External Validity, Journal of the Academy of Marketing Science 1999 = Bd. 27, S. 349–358.

Winteler, Adi & Peter Forster: Wer sagt, was gute Lehre ist? Evidenzbasiertes Lehren und Lernen, in: Brockmann/Dietrich/Pilniok, Methoden des Lernens in der Rechtswissen-schaft. Forschungsorientiert, problembasiert und fallbezogen, Baden-Baden 2012, S. 20–38.

Winter, Gerd: Tatsachenurteile im Prozess richterlicher Rechtssetzung, Rechtstheorie. Zeitschrift für Logik, Methodenlehre, Kybernetik und Soziologie des Rechts 1971 = Bd. 2, S. 171–192.

Wissenschaftsrat: Perspektiven der Rechtswissenschaft in Deutschland. Situation, Analysen, Empfehlungen, WR Drs. 2558-12 v. 9.11.2012, Zugriff am 12.12.2012 unter http://www.wissenschaftsrat.de/download/archiv/2558-12.pdf.

Witte, Erich: Köhler rediscovered: The anti-Ringelmann effect, European Journal of Social Psychology 1989 = Bd. 19, S. 147–154.

Wittenbaum, Gwen, Andrea Hollingshead & Isabel Botero: From cooperative to motivated information sharing in groups: moving beyond the hidden profile paradigm, Communication Monographs 2004 = Bd. 71, S. 286–310.

Wittenbaum, Gwen & Richard Moreland: Small-Group Research in Social Psychology: Topics and Trends over Time, Social and Personality Psychology Compass 2008 = Bd. 2, S. 187–203.

World Bank: GNI, Atlas method (current US$), The World Bank, Zugriff am 27.6.2014 unter http://data.worldbank.org/indicator/NY.GNP.ATLS.CD.

Woywode, Michael, Detlef Keese & Jan Tänzler: Corporate Governance in geschlossenen Gesellschaften – insbesondere in Familienunternehmen – unter besonderer Berücksichtigung von Aufsichtsgremien, Zeitschrift für Unternehmens- und Gesellschaftsrecht 2012 = Bd. 41, S. 418–445.

Wozner, Shai: Comment: Evidence-Based Law by Jeffrey J. Rachlinski, Cornell Law Review 2011 = Bd. 96, S. 925–930.

Yi, Young-Jong: Zur Strukturreform der koreanischen börsennotierten AG vor dem Hintergrund des deutschen und des US-amerikanischen Aktienrechts, Diss., Marburg 2004.

Yong, Ed: Replication studies: Bad copy, Nature 2012 = Bd. 485, S. 298–300.

Yong, Ed: Jede Menge Murks, Spektrum der Wissenschaft 2013, S. 58–63.

Young, Sam: Evidence-based management: a literature review, Journal of Nursing Management 2002 = Bd. 10, S. 145–151.

Yule, George Udny: An Introduction to the Theory of Statistics, London 1911.

Zajac, Edward J.: CEO Selection, Succession, Compensation and Firm Performance: A Theoretical Integration and Empirical Analysis, Strategic Management Journal 1990 = Bd. 11, S. 217–230.

Zarnoth, Paul & Janet Sniezek: The Social Influence of Confidence in Group Decision Making, Journal of Experimental Social Psychology 1997 = Bd. 33, S. 345–366.

Zedeck, Sheldon: An information processing model and approach to the study of motivation, Organizational Behavior and Human Performance 1977 = Bd. 18, S. 47–77.

Zeiler, Kathryn: Cautions on the Use of Economics Experiments in Law, Journal of Institutional and Theoretical Economics 2010 = Bd. 166, S. 178–193.

Zeisel, Hans: Empirische Rechtsforschung – ein natürlicher Zweig der Jurisprudenz, JuristenZeitung 1974 = Bd. 29, S. 561–564.

Zeisel, Hans & David Kaye: Prove It with Figures. Empirical Methods in Law and Litigation, New York 1997.

Ziliak, Stephen & Deirdre McCloskey: The Cult of Statistical Significance. How the Standard Error Costs Us Jobs, Justice, and Lives, Ann Arbor 2008.

Zimmerman, Emily: Do Grades Matter?, Seattle University Law Review 2012 = Bd. 35, S. 305–376.

Zweigert, Konrad & Hein Kötz: Einführung in die Rechtsvergleichung, 3. Aufl., Tübingen 1996.

Personenregister

Sachregister